

STIC-Biotech/Ch mLib

83753

From: Chan, Christina
Sent: Tuesday, January 07, 2003 10:25 AM
To: Duffy, Patricia; STIC-Biotech/ChemLib
Subject: RE: PLEASE RUSH SEQUENCE SEARCH--- Amendment due this Biweek
Importance: High

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Duffy, Patricia
Sent: Monday, January 06, 2003 10:43 AM
To: Chan, Christina
Subject: PLEASE RUSH SEQUENCE SEARCH--- Amendment due this Biweek
Importance: High

In re: 09/438,185

Please search SEQ ID NO:1047.
Please search na residues 1200675-1199590 of SEQ ID NO:1.
Please include an interference search.
Please print out top 50 hits in each category.

Thank you.

~~Patricia A. Duffy~~
~~CM1 6B06~~

AU 1645
703-305-7555

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

Searcher: _____
Name: _____
Position: _____
Date Picked Up: 1/7
Date Completed: 1/10
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 11:41:02 ; Search time 45 Seconds
(without alignments)
10621.161 Million cell updates/sec

Title: US-09-438-185a-1_COPY_1199590_1200675

Perfect score: 1086
Sequence: 1 ttggcaagtagctcaaac.....gttctctgcagtagtgac 1086

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	46.6	4.3	1642	10	US-09-205-658-312
C 2	41.4	3.8	817	10	US-09-205-658-314
C 3	40.6	3.7	640681	10	US-09-790-988-5
C 4	36.6	3.4	2098	10	US-09-971-388-5
C 5	36.4	3.4	1241	10	US-09-737-626A-25
C 6	36.4	3.4	2000	9	US-09-938-842A-3348
C 7	36.4	3.4	2004	9	US-09-687-576-256
C 8	36.4	3.4	326014	10	US-09-731-231A-3
C 9	36.2	3.3	580	10	US-09-920-300A-111
C 10	36.2	3.3	1985	10	US-09-817-913-4
C 11	36.2	3.3	1985	10	US-09-817-538-4
C 12	36.2	3.3	1985	10	US-09-764-869-1952
C 13	36.2	3.3	32191	10	US-09-764-869-1955
C 14	36.2	3.3	1128	10	US-09-815-242-4614
C 15	36	3.3	1134	10	US-09-815-242-8602
C 16	35.8	3.3	7311	12	US-10-044-090-455
C 17	35.8	3.3	7311	12	US-09-764-877-2599
C 18	35.4	3.3	952	10	US-09-764-877-2599
C 19	35.4	3.3	1134	10	US-09-815-242-8628

C 20	35.2	3.2	271	10	US-09-777-564-1653
C 21	34.8	3.2	335913	9	US-09-754-853A-2
C 22	34.8	3.2	335913	9	US-09-754-853A-3
C 23	34.6	3.2	1391	10	US-09-925-301-66
C 24	34.6	3.2	3092	10	US-09-925-302-310
C 25	34.4	3.2	10195	10	US-09-764-864-1600
C 26	34.2	3.1	4201	10	US-09-925-301-478
C 27	34	3.1	35641	10	US-09-962-436-306
C 28	33.8	3.1	180557	12	US-10-003-806-6
C 29	33.8	3.1	180557	12	US-10-003-806-9
C 30	33.6	3.1	147309	10	US-09-743-312-3
C 31	33.4	3.1	2679	10	US-09-880-107-2219
C 32	33.4	3.1	2680	10	US-09-782-378A-20
C 33	33.4	3.1	78036	9	US-10-109-551-1
C 34	33.4	3.1	1738	10	US-09-864-761-9781
C 35	33.2	3.1	2000	9	US-09-764-864-1784
C 36	33.2	3.1	198285	10	US-09-938-842A-4741
C 37	33.2	3.1	328	10	US-09-880-107-3814
C 38	32.8	3.0	406	10	US-09-878-574-3709
C 39	32.8	3.0	548	10	US-09-924-035A-596
C 40	32.8	3.0	806	10	US-09-939-980-6
C 41	32.8	3.0	31412	9	US-10-109-551-3
C 42	32.6	3.0	601	10	US-09-854-133-294
C 43	32.6	3.0	601	10	US-09-738-973-294
C 44	32.6	3.0	2000	9	US-09-938-842A-3797
C 45	32.6	3.0	303	10	US-09-969-708-291
C 46	32.4	3.0	1658	10	US-09-925-302-228
C 47	32.4	3.0	4436	9	US-09-522-334-26
C 48	32.4	3.0	513509	9	US-09-754-853A-4
C 49	32.4	3.0	23419	10	US-09-938-842A-4311
C 50	32.2	3.0	26006	10	US-09-764-869-1364
C 51	32.2	3.0	26006	10	US-09-764-869-1963
C 52	32.2	3.0	26006	10	US-09-764-864-1638
C 53	32.2	3.0	26013	10	US-09-764-869-1961
C 54	32.2	3.0	26013	10	US-09-764-864-1636
C 55	32.2	3.0	26018	10	US-09-764-869-1962
C 56	32.2	3.0	302350	10	US-09-764-864-1637
C 57	32.2	2.9	270	10	US-09-962-832-154
C 58	32.2	2.9	464	10	US-09-867-701-78
C 59	32.2	2.9	570	9	US-09-969-708-534
C 60	32.2	2.9	570	9	US-08-788-827-84
C 61	32.2	2.9	580	9	US-09-768-827-79
C 62	32.2	2.9	580	9	US-10-205-153-162
C 63	32.2	2.9	628	10	US-09-925-300-315
C 64	32.2	2.9	1079	10	US-09-947-971-3
C 65	32.2	2.9	1152	9	US-09-925-300-164
C 66	32.2	2.9	1152	9	US-10-012-896-997
C 67	32.2	2.9	1173	10	US-09-880-107-2297
C 68	32.2	2.9	1595	9	US-10-051-307-1
C 69	32.2	2.9	2000	9	US-09-938-842A-4033
C 70	32.2	2.9	2000	9	US-09-938-842A-4200
C 71	32.2	2.9	4857	10	US-09-764-847-1929
C 72	32.2	2.9	8220	10	US-10-153-273-11
C 73	32.2	2.9	8220	10	US-09-949-654-3
C 74	32.2	2.9	368004	9	US-10-002-344A-99
C 75	32.2	2.9	1253	9	US-09-764-869-2314
C 76	31.8	2.9	32249	10	US-09-983-965-2145
C 77	31.8	2.9	330	9	US-09-796-692-4292
C 78	31.6	2.9	375	9	US-09-970-966-41
C 79	31.6	2.9	396	10	US-09-825-294-41
C 80	31.6	2.9	492	10	US-09-954-456-46
C 81	31.6	2.9	492	10	US-09-954-456-962
C 82	31.6	2.9	495	10	US-09-783-590-2867
C 83	31.6	2.9	519	10	US-09-864-761-6821
C 84	31.6	2.9	1643	9	US-10-001-887-41
C 85	31.6	2.9	1643	9	US-09-989-919-21
C 86	31.6	2.9	1767	10	US-08-822-849A-349
C 87	31.6	2.9	2000	9	US-09-938-842A-4165
C 88	31.6	2.9	6823	10	US-09-954-456-1215
C 89	31.6	2.9	13957	10	US-09-782-378A-22
C 90	31.6	2.9	13957	10	US-09-880-107-2284
C 91	31.6	2.9	13957	10	US-09-880-107-2284
C 92	31.6	2.9	13957	10	US-09-880-107-2284

Sequence 1653, Ap
Sequence 2, Appli
Sequence 3, Appli
Sequence 66, Appl
Sequence 310, App
Sequence 1600, Ap
Sequence 478, App
Sequence 306, App
Sequence 2225, Ap
Sequence 6, Appli
Sequence 9, Appli
Sequence 3, Appli
Sequence 2219, Ap
Sequence 20, Appl
Sequence 1, Appli
Sequence 9781, Ap
Sequence 1784, Ap
Sequence 4741, Ap
Sequence 3814, Ap
Sequence 733, App
Sequence 3709, Ap
Sequence 396, App
Sequence 6, Appli
Sequence 3, Appli
Sequence 294, App
Sequence 294, App
Sequence 3797, Ap
Sequence 291, App
Sequence 228, App
Sequence 26, Appli
Sequence 4, Appli
Sequence 4311, Ap
Sequence 1364, Ap
Sequence 1663, Ap
Sequence 1638, Ap
Sequence 1961, Ap
Sequence 1636, Ap
Sequence 1662, Ap
Sequence 1637, Ap
Sequence 154, App
Sequence 78, Appl
Sequence 534, App
Sequence 84, Appl
Sequence 79, Appl
Sequence 162, App
Sequence 315, App
Sequence 3, Appli
Sequence 164, App
Sequence 997, App
Sequence 2297, Ap
Sequence 1, Appli
Sequence 4033, Ap
Sequence 4200, Ap
Sequence 1529, Ap
Sequence 11, Appl
Sequence 3, Appli
Sequence 99, Appl
Sequence 2314, Ap
Sequence 2145, Ap
Sequence 4292, Ap
Sequence 41, Appl
Sequence 41, Appl
Sequence 46, Appl
Sequence 962, App
Sequence 2867, Ap
Sequence 6821, Ap
Sequence 21, Appl
Sequence 349, App
Sequence 4165, Ap
Sequence 1215, Ap
Sequence 22, Appl
Sequence 2284, Ap


```

; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      3.7%; Score 40.6; DB 10; Length 640681;
Best Local Similarity 43.4%; Pred. No. 7.3;
Matches 187; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY   119 AAATGCTTATTCAAAATAAAGTCCTTTGGAGGTGATGTTGAAGGGAAGACGAATA 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190122 AAATATGATTTTTTATAAAAAAGTTGATAGTTTTAAAAAAAATAAAAAATGAACAAAAACA 190181

QY   179 ATCTGATCCAAATCTTAAGGGAGAAGACAGTGATATCAATGAAGCGTGCCAAGTTCC 238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190182 ATAGTGCTTAAAAAATAATAAAAAATAATTTTTCTCAATATAGCCTGTATTCCTGG 190241

QY   239 TGAGGAGAAGTGAATAAGAACGGCTCCCATATGCTTTCTTCCTTCATGGTTTCAATAAGT 298
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190242 AATCGAACAATCGTAAACAACAGCTGTAATAGATCTTTTAAATTTTTTAATATACAACAT 190301

QY   299 CCGTCTTCAACAGTAAACCACAAAGCAGGTCAATACGGATCAGATGCTTTGTAGGGTT 358
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190302 ACTAATTTAGTACAAAAAATAATCATATCATAAGAAAAATAATCATATTCGATACGCAT 190361

QY   359 TGTATGGCTGTGTTTTTACTAGGAAGAGCTTGATCTTTTCTATGACTTTTAGTGAAGAT 418
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190362 CATCTAAATTTAATAATAAACACACTCTTTAAATACITTTAGTAATAGTTTCTTTAGTCGT 190421

QY   419 CTGCCATGTTTATGAAAAATTCAGAAAAATGAGGAGATGTAGAACCCAGGCGTCCCT 478
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190422 GTAGATGTTTTTTTAAATGTTTTAAATTTGTAATGTGAATCGTAGTAAGTGAAGTATT 190481

QY   479 AAAAGGTCATGATGATGAGTACAGGAGTTAAGGAGAAATATATCTTTATCGAGAGTTCGAPT 538
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190482 ATATTAAACGGGAATATAAAAAATTTAAAAAATAATGCTCTTTATATAAAATTTTAAATT 190541

QY   539 ACAGAGGCAAT 549
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   190542 AAAGATCTAT 190552

RESULT 4
US-09-871-388-5/c
Sequence 5, Application US/09871388
Patent No. US20020127621A1
GENERAL INFORMATION:
APPLICANT: Rubin, Gerald M.
Pan, Duojia
Rooke, Jenny
Xavari, Reza
Yu, Tian
TITLE OF INVENTION: KUZ: A No. US20020127621A1 Family of Metalloproteases
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/871,388
FILING DATE: 31-May-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,931
FILING DATE: <Unknown>

```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2098 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-871-388-5

Query Match          3.4%; Score 36.4; DB 10; Length 2098;
Best Local Similarity 65.1%; Pred. No. 7;
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 816 AGCAACCTCCCTTTGATTTGATATGACAGCAAGATGACAAAGAGCTAAGTTTGA 875
DB 1136 AGCCACAAAAGCTTTGAAGTTTGTATATTCAGAGAAGATGACCAAGAGCTTCTTTATT 1077
QY 876 AAAATTCCTAAATAGAAATGCAT 898
DB 1076 AGGCTACTGAATATATTTAT 1054

RESULT 5
US-09-737-626A-25
; Sequence 25, Application US/09737626A
; Patent No. US20020144304A1
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen
; APPLICANT: Wilkins, Jack
; TITLE OF INVENTION: No. US20020144304A1 Plant Expression Constructs
; FILE REFERENCE: 38-21(51499)B
; CURRENT APPLICATION NUMBER: US/09/737,626A
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 09/737,626
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent version 3.0
; SEQ ID NO 25
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1241)
; OTHER INFORMATION: n= a or g or c or t/u
; OTHER INFORMATION: Act7 promoter polynucleotide sequence and intron
US-09-737-626A-25

Query Match          3.4%; Score 36.4; DB 10; Length 1241;
Best Local Similarity 59.8%; Pred. No. 6.3;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGAGTCT 420
DB 1030 TAGGTGTTGATGTTTACAAAGTGAATTCATGTTTCTCTTTGAGATCTGTGAAGTTGA 1089
QY 421 TCCCATGTTTATGAAAAATTCAGAAAAATGAGGATGTAGAAG 462
DB 1090 ACCTAGTTTTCATATATCAATATGACGCGATGTTTGAG 1131

RESULT 6
US-09-938-842A-3348
; Sequence 3348, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPL300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3348
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3348

Query Match          3.4%; Score 36.4; DB 9; Length 2000;
Best Local Similarity 59.8%; Pred. No. 7.8;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGAGTCT 420
DB 1789 TAGGTGTTGATGTTTACAAAGTGAATTCATGTTTCTCTTTGAGATCTGTGAAGTTGA 1848
QY 421 TCCCATGTTTATGAAAAATTCAGAAAAATGAGGATGTAGAAG 462
DB 1849 ACCTAGTTTTCATATATCAATATGACGCGATGTTTGAG 1890

RESULT 7
US-09-887-576-256
; Sequence 256, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256
; LENGTH: 2004
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-256

Query Match          3.4%; Score 36.4; DB 10; Length 2004;
Best Local Similarity 59.8%; Pred. No. 7.8;
Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 361 TATCGGTGTTTTTACTAGGAAGCTGTGACTTTTCTATGACTTTAGTGAAGAGTCT 420
```

[illegible]

QY	799	AACAGCCATATCAGTGGAGCAACCTTCCTTTGATTCTTTGTATATGACAGAGAATGAAC	858
Db	383	ATCAGGCCAATTTTTCGACCCTCTCTCATCTCAGAAATCTTGATCTTCATC	442
QY	859	AAGAGTAGTATTTTGAATAATTAAGAAATGCATTTGTCTGAGACTAAAGCTTG	918
Db	443	ACNAGCTATCGGCTGTGCTGATGCTCGAATAGAAAATCTCTTGCTGGATCTTCTCCATC	502
QY	919	CTTCTCTTATTTCTCTTTTGTAGA	943
Db	503	TTCATCTCCACTGCTTTCATGAACA	527
 RESULT 10 US-10-033-528-111 ; Sequence 111, Application US/10033528 ; Patent No. US20020131971A1 ; GENERAL INFORMATION: ; APPLICANT: King, Gordon E. ; APPLICANT: Meagher, Madeleine Joy ; APPLICANT: Xu, Jiangchun ; APPLICANT: Secrist, Heather ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY ; FILE REFERENCE: 210121.547C1 ; CURRENT APPLICATION NUMBER: US/10/033,528 ; CURRENT FILING DATE: 2001-12-26 ; NUMBER OF SEQ ID NOS: 1896 ; SOFTWARE: FastSeq for Windows Version 4.0 ; SEQ ID NO 111 ; LENGTH: 580 ; TYPE: DNA ; ORGANISM: Homo sapiens US-10-033-528-111			
 Query Match 3.3%; Score 36.2; DB 12; Length 580; Best local Similarity 53.1%; Pred. No. 5.2; Matches 77; Conservative 0; Mismatches 68; Indels 0; Gaps			
QY	799	AACAGCCATATCAGTGGAGCAACCTTCCTTTGATTCTTTGTATATGACAGAGAATGAAC	858
Db	383	ATCAGGCCAATTTTTCGACCCTCTCTCATCTCAGAAATCTTGATCTTCATC	442
QY	859	AAGAGTAGTATTTTGAATAATTAAGAAATGCATTTGTCTGAGACTAAAGCTTG	918
Db	443	ACNAGCTATCGGCTGTGCTGATGCTCGAATAGAAAATCTCTTGCTGGATCTTCTCCATC	502
QY	919	CTTCTCTTATTTCTCTTTTGTAGA	943
Db	503	TTCATCTCCACTGCTTTCATGAACA	527
 RESULT 11 US-09-817-913-4/c ; Sequence 4, Application US/09817913 ; Patent No. US20020061860A1 ; GENERAL INFORMATION: ; APPLICANT: Li, Zuomei ; APPLICANT: Bonfils, Claire ; APPLICANT: Besterman, Jeffrey ; TITLE OF INVENTION: Inhibition of Specific Histone Deacetylase Isoforms ; FILE REFERENCE: 106101.145 ; CURRENT APPLICATION NUMBER: US/09/817,913 ; CURRENT FILING DATE: 2001-03-26 ; PRIOR FILING DATE: 2000-03-26 ; PRIOR APPLICATION NUMBER: US 60/192,157 ; NUMBER OF SEQ ID NOS: 33 ; SOFTWARE: PatentIn version 3.1 ; SEQ ID NO 4 ; LENGTH: 1985 ; TYPE: DNA ; ORGANISM: Human			


```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4614
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
;
US-09-815-242-8602

Query Match          3.3%  Score 36;  DB 10;  Length 1128;
Best Local Similarity 50.0%; Pred. No. 7.7;  Mismatches 0; Gaps 0;
Matches 90; Conservative 0;

Qy 35 TCTTGATTGTAAGGGGAATGATCTACAGACACCTTGTCGAGCATCCATCTTAATTTT 94
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 832 TCTTTATTAGCAGCGTGAAGGTTGTAATACGTTTAAATTAAGTACAGGTGATGCGTT 891
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 95 GAAGTGAGTTCTACAGTTCAATCAAAATGCTTATTGAAATAAAGTCTCTTGAGGTT 154
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 892 GTTGAATTAATCTTACATACACCAAGAAATGCTGTAAGAAAGAAAGTGTGATGCAAC 951
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 155 GATGTTATGAAGGGAAGACGAATAATCTGATCAATTTCTAAAGGGAGAACACGTACGTTA 214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 952 GATGTTGAAGGTGGTAGCTGAAATTTTCATCACTCTAAATATATGATGATGTCCTTA 1011

RESULT 16
; Sequence 8602, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8602
```

```

; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1).....(1134)
US-09-815-242-8602

Query Match          3.3%  Score 36;  DB 10;  Length 1134;
Best Local Similarity 50.0%; Pred. No. 7.7;  Mismatches 0; Gaps 0;
Matches 90; Conservative 0;

Qy 35 TCTTGATTGTAAGGGGAATGATCTACAGACACCTTGTCGAGCATCCATCTTAATTTT 94
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 TCTTTATTAGCAGCGTGAAGGTTGTAATACGTTTAAATTAAGTACAGGTGATGACGTT 894
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 95 GAAGTGAGTTCTACAGTTCAATCAAAATGCTTATTGAAATAAAGTCTCTTGAGGTT 154
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 895 GTTGAATTAATCTTACATACACCAAGAAATGCTGTAAGAAAGAAAGTGTGATGCAAC 954
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 155 GATGTTATGAAGGGAAGACGAATAATCTGATCAATTTCTAAAGGGAGAACACGTACGTTA 214
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 GATGTTGAAGGTGGTAGCTGAAATTTTCATCACTCTAAATATATGATGATGTCCTTA 1014

RESULT 17
US-10-044-090-455
; Sequence 455, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 455
; LENGTH: 7311
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No. US20020137081A1 347926.5
; NAME/KEY: unsure
; LOCATION: 5438, 5455, 5470, 5482, 5495, 5512, 5559
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-455

Query Match          3.3%  Score 35.8;  DB 12;  Length 7311;
Best Local Similarity 50.9%; Pred. No. 19;  Mismatches 82; Indels 0; Gaps 0;
Matches 85; Conservative 0;

Qy 792 TGGTTACACACCCCATATCATGTCGAGACACCTTCCCTTTGATTCCTTTGATTAATGACAGAA 851
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2391 TGATTACTGTAGTCAATATTATGAAAAAGGTTTGTGTTTACTCTCTAGTAGAGAAA 2450
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 852 GATGAACAAAGAGTAAGTTTGTAAAAATTTCTAAAAATAGAAATGCAATTTGTGCGAGCTA 911
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2451 GTGGGCAAAATATATCTTTTGAATAAATGCTATATGCGACCTAAATTTTCTTTT 2510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 912 AAGCTTGCCTTCTTTTATTTTCCCTTTTGTAGAATGATCGGTAGTAG 958
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2511 AAAATGCCITTAAGTTGCAGTCTCATTTTGTATAATCATTTTGTCTCCAG 2557
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 18
US-09-764-877-2599/C
; Sequence 2599, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; SEQ ID NO 8602
```

```
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2599
; LENGTH: 952
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2599

Query Match      3.3%; Score 35.4; DB 10; Length 952;
Best Local Similarity 47.9%; Pred. No. 10;
Matches 102; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 712 AAATCTGGACAATAGCTTTTCCATAGAGAAACCGTGGAGGAAGAGTGCATACCATAA 771
DB 259 ATATATAGGAAACTCCCTTTAAATAAGTGTAAATTCATTAAGATGAAGTGAATAAAGA 200

QY 772 ATTCGGATTAAAAAGGGAGTCTTTACAAAGCCATATCATGTGGAGCAACCTTCCCTTTG 831
DB 199 TGACAGAAATATAGAACATCTAATTTTCAATGCAAAATCATATGTGAAAAAATGAACACAT 140

QY 832 ATTCCTTTGTAATGACAGAGATGAACAAAGAGTAAGTTTTTTGAAAAAATTCATAAATAGA 891
DB 139 AATTTAAATAAATAATAGGATAGTAGATGACCATGTCAACTAAATCTCTTAAATAACA 80

QY 892 AATGCATTTTGTGCGAGCTAAAGCTTGTCTCTC 924
DB 79 AAAGTCTCACTCATATATAAACTTATACCTC 47

RESULT 19
US-09-815-242-8628
; Sequence 8628, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8628
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (1)...(1134)
US-09-815-242-8628

Query Match      3.3%; Score 35.4; DB 10; Length 1134;
Best Local Similarity 49.7%; Pred. No. 11;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 34 CTCTTTGATTCTAAAGGGGAATTGATTTCAACAGACACCTTGGTCGAGCATCATTTCTAATTT 93
DB 834 CTCCTTTATTAGCGCGTGAAGGTGGTAATAACGTTTATTAATTAAGTACAGGTGATGCGT 893

QY 94 TGAAGTCAGATTCTACCAAGTTCATCAAAATGCTTTATTGAAAAATAAAGTCTCTTGTGGAGT 153
DB 894 TGTGAATTCCTCTTCATCAATCAGCAAAATTTGGTACTGTAAAGAAAGAAAGTTGATGCAAA 953

QY 154 TGATGTATTGAAGGGGAAGACGAATTAATCTGATCCCAATTTCTAAAGGGGAGAACACGTTACGTT 213
DB 954 CGATGTTGAAGGTGGTAGCCTGAAATTTTCAATCAACTCTAAATATATATGATGATGCTTT 1013

QY 214 A 214
DB 1014 A 1014

RESULT 20
US-09-777-564-1653/c
; Sequence 1653, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1653
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(271)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-1653

Query Match      3.2%; Score 35.2; DB 10; Length 271;
Best Local Similarity 52.4%; Pred. No. 6.8;
Matches 76; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 799 AACAGCCATATCATGTGGAGCAACCTTCCCTTTGATTCTTTGATAATGACAGAGATGAAC 858
DB 248 ATCAGCCACNTTCTTCGACCTCTCTTCCTTCATCTCAGAAATCTGAGAATCTTTCATC 189

QY 859 AAAGAGTAAGTTTTTGTAAAAAATCTTAAATAGAAATGCATTTGTGTGAGCTAAAAGCTTG 918
DB 188 ACAAGCTATCCGCTTGTCTGTGATGCTCGAATAGAAATTTCTCTGTGTGATCTTCTCCATC 129

QY 919 CTCTCTCTTATTCTTCTCTTTTGTAGA 943
DB 128 TTCATCTCCACTGTCTTTCATGAACA 104

RESULT 21
US-09-754-853A-2
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
```



```

; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754, 853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174, 880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
; US-09-754-853A-2

```

Query Match	3.28;	Score 34.8;	DB 9;	Length 335913;
Best Local Similarity	44.9%;	Prod. No. 1.6e-02;		
Matches 132;	Conservative	Mismatches 162;	Indels 0;	Gaps 0;
QY 574	AGACAAGTATTGTATGGGAGCTACAAATCTCGAAACGGGATAATNAGGAAAATGTTGTTTC	633		
Db 199825	AAACGACGATGTCTTTTGTATPAGACAGTGTGCAATTCGAAGATCAACAGGTTCTTTGACCC	199884		
QY 634	TAAATTCGAAGAAATTTAATGACTGCTTGATGCTTAAGAAATCAGAAAGGAGACCGGAAAGC	693		
Db 199885	TTAACCTTTGCATTGCTCATCATCATGCTGCGAAAATCAACGACTTAAAGAAACGGTATGCG	199944		
QY 694	TTCTTAAGTAACTCAAGAAAAAATCTTTGGACAATAGCTTTTCCATAGAGAAAACCGTGAGGA	753		
Db 199945	ATTTTAGAATGTAAAAAAATTTGGTGCTGTCTATATCATTAATTAATTAAGGAAAAG	200004		
QY 754	NAAGGAGTCGATACCATATAATTCGGATTAAAAAGGGAGTGCTTACAAACGACCATATCAGT	813		
Db 200005	AGTGTCAGGTCACACAAATTTTGAACTTTTTCATTTTATTTAATTTGGCATTTTGTGTAT	200064		
QY 814	GGAGCAACCTTCCTTTTGATCTTTTGATAATGACAGAAAGATCAACAAAGAGATAA	867		
Db 200065	TTTGAAATCTCTCTATTTACTCTCTTAATGAGCTTTTGACHTTTAATTTATGATTAA	200118		

```

RESULT 22
US-09-754-853A-3
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754.853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763), (48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-3

```

Query Match	3.2%;	Score	34.8;	DB	9;	Length	335913;
Best Local Similarity	44.9%;	Pred.	No.	1.6e+02;			

	Matches	132;	Conservative	0;	Mismatches	162;	Indels	0;	Gaps	0;
Qy	574	AGACAAGTATTGATGGGAGCTACAAATCCTGAACGGGATTAATAGAGAAAAATCTCTTTTC	633							
Db	199825	AAACGACGATTCCTTTTCTATAGGACAGTTCGCATTAAGATTAACAGGTTCTTTGACC	199884							
Qy	634	TAATTCGAAGATTTTAATGACTGCTTGATGGCTTAGAAATCAGAAAGAGACCGAAACG	693							
Db	199885	TTAACCTTTTCGATTCGTCATCATCATCATCGTGGAAAAATCAAGCACTTAAGAAACCGTATGC	199944							
Qy	694	TTCTAAGTAATCAAGAAAAAATCTTGGACAATAGCTTTTTCATAGAGAAACCGTGAGGA	753							
Db	199945	ATTTTGAAGATGTAAAAAATTTGGTGGTCTCTATATCAATTAATCACTTAACTTAAAGGAAAG	200004							
Qy	754	AAGGAGTCGATACCATTAATCCGATTAAAAAGGGAGTGCTTTACACAGCCATATCACT	813							
Db	200005	AGTGATACAGGTACACAAATTTTGAACTTTTTCATTTTATTTAATTTGGCATTTTGCTAT	200064							
Qy	814	GGAGCAACCTTCCTTTTGATTCCTTTGATTAATGACAGAAGATGAACAAAGAGTAA	867							
Db	200065	TTTCAGATCTCTTTATTTAGTCCCTTAATGAGCTTTTGAACCTTTTAATTAATGATTA	200118							

```

RESULT 23
US-09-925-30166
; Sequence 66, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05682
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 1391

```

```

1 ORGANISM: Homo sapiens
2
3 FEATURE:
4 NAME/KEY: misc_feature
5 LOCATION: (1)
6 OTHER INFORMATION: n equals a,t,g, or c
7
8 NAME/KEY: misc_feature
9 LOCATION: (15)
10 OTHER INFORMATION: n equals a,t,g, or c
11
12 NAME/KEY: misc_feature
13 LOCATION: (20)
14 OTHER INFORMATION: n equals a,t,g, or c
15
16 NAME/KEY: misc_feature
17 LOCATION: (25)
18 OTHER INFORMATION: n equals a,t,g, or c
19
20 NAME/KEY: misc_feature
21 LOCATION: (27)
22 OTHER INFORMATION: n equals a,t,g, or c
23
24 NAME/KEY: misc_feature
25 LOCATION: (1343)
26 OTHER INFORMATION: n equals a,t,g, or c
27
28 NAME/KEY: misc_feature
29 LOCATION: (1358)
30 OTHER INFORMATION: n equals a,t,g, or c
31
32 US-09-925-301-66

```

```

Query Match          3.2%; Score 34.6; DB 10; Length 1391;
Best Local Similarity 57.0%; Pred. No. 19;
Matches 61; Conservative 0; Mismatches 46; Indels 0; Gaps
QY 628 TGTGTTCTAATTCGAACAATTTAATGACTGTGTTGATGTCTCTAGAAGATCAGAAAGGAGACC 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 26358 TACTTTCAAAGTGTAATACATCGTTGTACTTATCTGTCAGTGAGAAAAAGATTAAAA 26417

Qy 861 AGAGTAAGTTTT 873

Db 26418 ACAGAAGTTTT 26430

RESULT 31

US-09-742-312-3

; Sequence 3, Application US/09742312

; Patent No. US20020045166A1

; GENERAL INFORMATION:

; APPLICANT: CHANDRAMOULISARAN, Ishwar et al

; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,

; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: CL000838

; CURRENT APPLICATION NUMBER: US/09/742.312

; CURRENT FILING DATE: 2000-12-22

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 147309

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(147309)

; OTHER INFORMATION: n = A,T,C or G

US-09-742-312-3

Query Match 3.1%; Score 33.6; DB 10; Length 147309;
Best Local Similarity 48.0%; Pred. No. 2.4e+02;
Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 694 TTCTAAGTAATCAAGAAAAATCTTTGGACAATAGCTTTTCCATAGAGAAAAACCGGTGAGGA 753

Db 35134 TTCAAGTCAACAATCCATATTTTTCAGACAACATGTTGTCTATGTAGAAAAATCCCAAAG 35193

Qy 754 AAGGAGTCGATACATAAATCCGATTTAAAAAGGGAGTGCTTACACAGCCATATCACT 813

Db 35194 AATCAAGGCACCTAGTAACTAGTTCAGTTAGCAAGTGCACAGATATAAGGACACACACAA 35253

Qy 814 GGAGCAACCTCCCTTTGATCTTTTGATTAATGACAGAGATGAACAAAGTAAGTTTTT 873

Db 35254 GAATCAACATTTCTATATATATATGATGAATTAATGAAGAAAAATTTGTAATATTT 35313

Qy 874 GAAAAATTCATAATAGAA 893

Db 35314 CCAATAGCTGAAAAAATGAA 35333

RESULT 32

US-09-880-107-2219/c

; Sequence 2219, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; APPLICANT: Gene Logic, Inc.

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880.107

; CURRENT FILING DATE: 2001-05-14

; PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-06-14

; PRIOR FILING DATE: 2000-06-14

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2219

; LENGTH: 2679

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L47726

US-09-880-107-2219

Query Match 3.1%; Score 33.4; DB 10; Length 2679;

Best Local Similarity 52.5%; Pred. No. 52;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 180 TCTGATCCAAATTTCTAAAGGGAGAACACAGTACCTTATCAATGAAAGCGGTGCCAAGTTCTT 239

Db 1586 TCTTCTCCAGCTCCAGGGGAGAAAGCTTTGGCTTCTCTGATAAGCAGTACTGTAATTTCAC 1527

Qy 240 GAGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGTTTTCATTAAGTC 299

Db 1526 CAAAGGATGACAGAGGCCACCATATGCTTTTATGGAGTCTCTTGTTCGAGAGCC 1467

Qy 300 CGCTTTTCAACAGTAAACCA 318

Db 1456 CAAACTCCACAGTAAACCA 1448

RESULT 33

US-09-782-378A-20/G

; Sequence 20, Application US/09782378A

; Patent No. US20020102751A1

; GENERAL INFORMATION:

; APPLICANT: Hearing, Patrick

; APPLICANT: Bahou, Wadie

; APPLICANT: Sandalon, Ziv

; APPLICANT: Gnatenko, Dmitri

; TITLE OF INVENTION: Adenoviral Vectors

; FILE REFERENCE: STONYB-04970

; CURRENT APPLICATION NUMBER: US/09/782.378A

; CURRENT FILING DATE: 2001-02-12

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 27

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 20

; LENGTH: 2680

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-782-378A-20

Query Match 3.1%; Score 33.4; DB 10; Length 2680;

Best Local Similarity 52.5%; Pred. No. 52;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 180 TCTGATCCAAATTTCTAAAGGGAGAACACAGTACCTTATCAATGAAAGCGGTGCCAAGTTCTT 239

Db 1586 TCTTCTCCAGCTCCAGGGGAGAAAGCTTTGGCTTCTCTGATAAGCAGTACTGTAATTTCAC 1527

Qy 240 GAGGAGAACTGATAAGAACGGCTCCATATGCTTTTCTCTTCATGTTTTCATTAAGTC 299

Db 1526 CAAAGGATGACAGAGGCCACCATATGCTTTTATGGAGTCTCTTGTTCGAGAGCC 1467

Qy 300 CGCTTTTCAACAGTAAACCA 318

Db 1456 CAAACTCCACAGTAAACCA 1448

RESULT 34

US-10-109-551-1

; Sequence 1, Application US/10109551

; Publication No. US20020194635A1

; GENERAL INFORMATION:

; APPLICANT: DUNNE, PATRICK W.

; APPLICANT: PIEDRAHITA, JORGE

; TITLE OF INVENTION: TRANSGENIC ANIMALS RESISTANT TO TRANSMISSIBLE

; TITLE OF INVENTION: SPONGIFORM ENCEPHALOPATHIES

; FILE REFERENCE: TAMK:20705

RESULT 37
 US-09-938-842A-4741
 ; Sequence 4741, Application US/09938842A
 ; Patent No. US20020160378A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Harper, Jeff
 ; APPLICANT: Kreps, Joel
 ; APPLICANT: Wang, Xun
 ; APPLICANT: Zhu, Tong
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 ; FILE OF INVENTION: SAME, AND METHODS OF USE
 ; FILE REFERENCE: SCRI1300-3
 ; CURRENT APPLICATION NUMBER: US/09/938,842A
 ; CURRENT FILING DATE: 2001-08-24
 ; PRIOR APPLICATION NUMBER: US 60/227,866
 ; PRIOR FILING DATE: 2000-08-24
 ; PRIOR APPLICATION NUMBER: US 60/264,647
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/300,111
 ; PRIOR FILING DATE: 2001-06-22
 ; NUMBER OF SEQ ID NOS: 5379
 ; SEQ ID NO 4741
 ; LENGTH: 2000
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-09-938-842A-4741

 Query Match 3.1%; Score 33.2; DB 9; Length 2000;
 Best Local Similarity 49.4%; Pred. No. 51;
 Matches 86; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

 Qy 275 CTTCCCTTCATGGTTTCAATAAGTCGCTTTCAACAGTAAACCAAGCAGCGTACAAATA 334
 Db 356 CATAATTCACCAATTCAGAGAAAAGCTTTTAAATCATGTGATCTAATAAGGTTTACAAAT 415
 Qy 335 GCGATCAGATTCCTGTAGGTTTGTATGGTTTCTTCTAGTACAGAGAGCTTGTACT 394
 Db 416 GTTCGACCTACATTTTGGAGAAGTATTCGTCATCTATTAGTAGTACTGATAT 475
 Qy 395 TTTTCTATGACATTTAGTAGAGAGCTTCCCATGTTTATGAAATTCAGAAAT 448
 Db 476 GTTGGCTTTCTTTCTCGAAGATCAGATATCGCCACAAATATTTTGTATAT 529

 RESULT 38
 US-09-880-107-3814
 ; Sequence 3814, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3814
 ; LENGTH: 198285
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X87344
 US-09-880-107-3814

 Query Match 3.1%; Score 33.2; DB 10; Length 198285;
 Best Local Similarity 48.4%; Pred. No. 3.3e+02;

Matches 92; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

 Qy 698 AAGTAATCAAGAAAAAATCTTGACAATAGCTTTTCCATAGAGAAAACCGTGAGAAAGG 757
 Db 149825 AAATAATTAATAAATAGATGGACTTCTAGCTAGACTAATAAAGAGAAAAGAGAGAGG 149884

 Qy 758 AGTCGATACCAATAAATTCGGATTAAAAAAGGGAGTGCTTCAACAGCCATATCATGTGGAG 817
 Db 149885 ATTCAATAAACACAAATTAGGAATGACAAGGAGGTATTACCACTGACCCACAGAAATA 149944

 Qy 818 CAACCTTCCCTTTGATTCTTGTATATGACAGAGATGAACAAGAGTAAGTTTTCGAAA 877
 Db 149945 CAGACAACCAATCAGACAATATTATGTAATCTTCTATGACATGAAGTAGAAAACTAGAAG 150004

 Qy 878 AATCTAAAA 887
 Db 150005 AATGGATAA 150014

 RESULT 39
 US-09-878-574-733
 ; Sequence 733, Application US/09878574
 ; Patent No. US20020110548A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrum, Joseph R.
 ; APPLICANT: Thompson, Michael D.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(15401)B
 ; CURRENT APPLICATION NUMBER: US/09/878,574
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/333,535
 ; PRIOR FILING DATE: 1999-06-14
 ; NUMBER OF SEQ ID NOS: 15775
 ; SEQ ID NO 733
 ; LENGTH: 328
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; OTHER INFORMATION: Clone ID: LIB3028-047-Q1-B1-A3
 US-09-878-574-733

 Query Match 3.0%; Score 32.8; DB 10; Length 328;
 Best Local Similarity 58.0%; Pred. No. 30;
 Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

 Qy 837 TTGATAATGACAGAGATGAACAAAGAGTAAGTTTGTGAAAATTTCTAAATAGAAATGC 896
 Db 213 TTGATAATCTCTGAAGTTGAAGTATGACAGCGTTTGTGAGATGGCTTTCTAATAGACTAC 272

 Qy 897 ATTTGTGTCGAGCTAAAGCTTGTCTTCTTTTATTTTCCTT 936
 Db 273 ACTTTTGTGCTAATACAGTTGATGCGCTGCTTCTTCTT 312

 RESULT 40
 US-09-878-574-3709/C
 ; Sequence 3709, Application US/09878574
 ; Patent No. US20020110548A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Byrum, Joseph R.
 ; APPLICANT: Thompson, Michael D.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; FILE REFERENCE: 38-21(15401)B
 ; CURRENT APPLICATION NUMBER: US/09/878,574
 ; CURRENT FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 09/333,535
 ; PRIOR FILING DATE: 1999-06-14
 ; NUMBER OF SEQ ID NOS: 15775
 ; SEQ ID NO 3709
 ; LENGTH: 406

TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-006-Q1-B1-H7
US-09-878-574-3709

Query Match 3.0%; Score 32.8; DB 10; Length 406;
Best Local Similarity 48.0%; Pred. No. 33;
Matches 94; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 517 ATCTTTATCGAGAGTTCGCAATACAGAGCCCAATGGGAAATACGCTCTGCAACAGAGA 576
DB 274 ATCTCTACCCGCCATGAGCAATGCCATGCAACCGGAAGCTAAAGGAATAATATGTCAGT 215
QY 577 CAAATGATTGATGGGAGCTACAAATCTTGAACGGGATAATAGGAAAAATGTTTCTAA 636
DB 214 CTTGTACTGACAAATAGCTGATGAGTGTAAATGTGTTATGACAGGCTTTTCTCC 155
QY 637 TTCGAGAAATTTATGACTGCTTGATGCTTAAGAAATCAGAAAGGAGACCGAAAGCTTC 696
DB 154 TTCAGTGTGTGATCAGATCTATCTGCTGCTGACGAGTGTAAATGTTGATGACAGGCTTTTCTCC 155
QY 697 TAAGTAATCAGAAAA 712
DB 94 AAACAAATCAAGAA 79

RESULT 41
US-09-924-035A-396/c
Sequence 396, Application US/09924035A
Patent No. US20020142319A1
GENERAL INFORMATION:
APPLICANT: Glaxo, Jrn
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2011US
CURRENT APPLICATION NUMBER: US/09/924,035A
CURRENT FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/148,784
PRIOR FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 900
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 396
LENGTH: 548
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(548)
OTHER INFORMATION: n = A,T,C or G
US-09-924-035A-396

Query Match 3.0%; Score 32.8; DB 10; Length 548;
Best Local Similarity 48.9%; Pred. No. 38;
Matches 88; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 350 TGTAGGTTTGTATCGCTGTTGTTTACTAGGAAGAGCTGTGACTTTTCTATGACTTTA 409
DB 189 TTTGAGATTGTTGTTCTTACTTCTTACTTGTAGTTGTATACACATTTTGTGAAGACCTT 130
QY 410 GTGAAGAGTCTCCCATGTTTATGAAAAATTCAGAAATCAGGGGATGTAGAGCCAAAGC 469
DB 129 GAGAATGATCATCAGTTCTCTGATTAAACGCGAGAGATTATATAATTTGGGCCCAAG 70
QY 470 ACGTGCCCTAAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 529
DB 69 TGTAGGCCCATAGATTATCAAAAAAGCCGCCGAATTCGATATCAAGCTTATCGATA 10

RESULT 42
US-09-939-980-6
Sequence 6, Application US/09939980
Patent No. US20020082234A1
GENERAL INFORMATION:
APPLICANT: DUNNE, PATRICK W.

APPLICANT: Black, Michael
Burnham, Martin
Hodgson, John
Knowles, David
Lonetto, Michael
Nicholas, Richard
Pratt, Julie
Reichard, Richard
Rosenberg, Martin
Ward, Judith
TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,980
FILING DATE: 27-Aug-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/936,165
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 806 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-939-980-6

Query Match 3.0%; Score 32.8; DB 10; Length 806;
Best Local Similarity 53.6%; Pred. No. 44;
Matches 67; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 27 GATATTCCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 86
DB 190 GAGACCCCTAATCACTGAGGAGCTGATTCATATATCGAAATATGATGATGATGATGATGAT 249
QY 87 CTAAATTTGAAGTGAAGTTCACGATTCATCAAAATGCTTTATTGAAAAATAAATCTCTT 146
DB 250 CATAATTTGAAGAAATTTCTATCAGAGGACAAATTTGTTAAAGCTAAATTTGTCTATA 309
QY 147 GTGGA 151
DB 310 GATGA 314

RESULT 43
US-10-109-551-3/C
Sequence 3, Application US/10109551
Publication No. US20020194635A1
GENERAL INFORMATION:
APPLICANT: DUNNE, PATRICK W.

Query	Subject	Score	DB	Length	Mismatches	Indels	Gaps
124	TCCTATTGAAATAAAGTCTCTGTGGATGTCATGATTGAAGGAGCAGCAATAATCTG	183					
702	TCAGATAGTAAATAGTTACTCAAAATACGTTAACATCGAGAAAAAATGAAAAAATCTG	761					
184	ATCCAATTCCTAAAGGAGACAGCTAGTTCATCAATGAAAGCGTGTCCAAGTTCCTGAGG	243					
762	TAAGCCATTTATCATACAACAATATTTTTTCTTTTAAAGTAGCTGGTTTCTTCATG	821					
244	AGAAGCTGATAAGACGGCTCCATATGCTTTCTCTTCATGTTTCAATAAGTCGCT	303					
822	CTTGATATTAAGATTAACCAATATCATCTCTTTCTTTCTTTTCAATTAATATCAT	881					
304	TTCAACAGTAAACCAAGACGCGTCATACATAGCGA	338					
882	TAAATAAAGAGAAAGACGCTGATACATATACCA	916					
<p>Query Match 3.0%; Score 32.6; DB 9; Length 2000;</p> <p>Best Local Similarity 47.0%; Pred. No. 73;</p> <p>Matches 101; Conservative 0; Mismatches 114; Indels 0; Gaps 0;</p>							
124	TCCTATTGAAATAAAGTCTCTGTGGATGTCATGATTGAAGGAGCAGCAATAATCTG	183					
702	TCAGATAGTAAATAGTTACTCAAAATACGTTAACATCGAGAAAAAATGAAAAAATCTG	761					
184	ATCCAATTCCTAAAGGAGACAGCTAGTTCATCAATGAAAGCGTGTCCAAGTTCCTGAGG	243					
762	TAAGCCATTTATCATACAACAATATTTTTTCTTTTAAAGTAGCTGGTTTCTTCATG	821					
244	AGAAGCTGATAAGACGGCTCCATATGCTTTCTCTTCATGTTTCAATAAGTCGCT	303					
822	CTTGATATTAAGATTAACCAATATCATCTCTTTCTTTCTTTTCAATTAATATCAT	881					
304	TTCAACAGTAAACCAAGACGCGTCATACATAGCGA	338					
882	TAAATAAAGAGAAAGACGCTGATACATATACCA	916					
<p>Query Match 3.0%; Score 32.4; DB 10; Length 303;</p> <p>Best Local Similarity 49.0%; Pred. No. 37;</p> <p>Matches 75; Conservative 0; Mismatches 78; Indels 0; Gaps 0;</p>							
37	TTGATGTTGAAGGGGAATTGATTCCTAACAGACCTTGGTCGAGCATCCATTCATTTTGA	96					
1	TTCCNNTAACATGTAGATTTATTTTAACTCAGTTTGGTACATACAGATATGGTTTTCG	60					
97	AGTGAGTTCTACCAAGTTCATCAAAAATGCTTTATTGAAATAAAGTCTCTTGTGGAGTTGA	156					
61	NGTTTTTAATCAACTGANATAGAAATGCTTCTNACTNCGCAGTATCTGCTGTGCAACAA	120					

[illegible]

US-09-522-334-26

[illegible]

RESULT 50

```

US-09-754-853A-4
; Sequence 4, Application US/09754853A
; Publication NO. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 4
; LENGTH: 513509
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (111805)..(113968),(114684)..(115204)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(513509)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 3180I3_region_A3
US-09-754-853A-4

```

Query Match	3.08;	Score 32.4;	DB 9;	Length 513509;
Best Local Similarity	53.1%;	Pred. NO. 5.3e+02;		
Matches 69;	Conservative	0;	Mismatches 61;	Indels 0;
				Gaps 0;
QY 625	ATGTGTTCTCTAATTCGAGAATTTTAATGACGTGCTTCATGGTCTTAAGAAATCAGAAAAGGAG	684		
Db 263074	ATGAAAATATAAAATTTTGAATATATATTTTATGTTTGTTCACGAATAAAAACGAAT	263133		
QY 685	ACCGAAAGCTTCFAAGTATCAAGAAAAAATCTTGGACAATAGCTTTTCCATAGAGAAAA	744		
Db 263134	TTTTAAATGTTCTATGAATTTTGAAAAAATATCTTTTAGAGAAAAATATTTTATATAATG	263193		
QY 745	CCGTGAGGAA	754		
Db 263194	AATTGACTAA	263203		

Search completed: January 9, 2003, 15:15:20
Job time : 2105 secs

Db 908 TAAACTAAGTGTGTGTATGAATGCTTTAAGGAGGCTTCCTTTCTTAACAGATTGG 849
QY 1012 GGAAG 1017
Db 848 GTGAGG 843

RESULT 2
US-09-813-133A-3/C
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weinliu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813/133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match 3.7%; Score 39.8; DB 4; Length 55827;
Best Local Similarity 48.1%; Pred. No. 0.46; 122; Indels 0; Gaps 0;
Matches 113; Conservative 0; Mismatches 0;

QY 522 TATCGAGAGTTCCGATTACAGACCAATGGGAAAGTACGGTCTCGCAACAGACAACT 581
Db 29992 TCTCAAAAGCAGCTCAACAGACCAAAATAGATAAAGGGGATTACATCAAACTGAAA 29933
QY 582 ATTGATGGGGAGCTACAAATCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGA 641
Db 29932 TTTCCTTCGACACGAAAGGAAATACACGCTGAAAGAAAAAATGGGAGAAATATG 29873
QY 642 AGAATTTAATGACTGTGTGTTGTTGTTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAAGT 701
Db 29872 TGCAAGTCTCAGCAACAGGGGACTTAATACCAAGATATATAAGGACGCAAAACAATT 29813
QY 702 AATCAAGAAAAATCTTGAGCAATAGCTTTTCATAGAGAAAAACCGTGAGGAAAG 756
Db 29812 CAAAGAAAAAACAATAATCAATTTTAAAAATGAGAAAGGAATCTGAACAG 29758

RESULT 3
US-08-937-931-5/C
; Sequence 5, Application US/08937931
; Patent No. 5935792
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 5935792el Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/285,502
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2098 base pairs
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/937,931
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: B97-081
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2098 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-937-931-5

Query Match 3.4%; Score 36.6; DB 2; Length 2098;
Best Local Similarity 65.1%; Pred. No. 0.87; 29; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 0;

QY 816 AGCAACCTTCCCTTTGATTCTTTGATAATGACAGAAGATGAACAAAGATAGTTTGA 875
Db 1136 AGCCACAAAAGCTTGAAGGTTTGATATTCAGAGAAGATGACCAAAAGAGTTCTTTTATTT 1077

QY 876 AAAATCTTAAATAGAAATGCAT 898
Db 1076 AGGCTACTGAATATTATTATTTAV 1054

RESULT 4
US-09-285-502-5/C
; Sequence 5, Application US/09285502
; Patent No. 6190876
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerald M.
; APPLICANT: Pan, Duojia
; APPLICANT: Rooke, Jenny
; APPLICANT: Yavari, Reza
; APPLICANT: Xu, Tian
; TITLE OF INVENTION: KUZ: A No. 6190876el Family of Metalloproteases
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/09/285,502
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/937,931
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B97-081
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2098 base pairs
; CURRENT APPLICATION DATA:

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-285-502-5

Query Match 3.4%; Score 36.6; DB 4; Length 2098;

Best Local Similarity 65.1%; Pred. No. 0.87; Mismatches 0; Indels 0; Gaps 0;

Matches 54; Conservative

OY 816 AGCAACCTCCCTTGGATGATGACAGCAAGATGACAAAGAGTAAAGTTTGA 875

DB 1136 AGCCAAAGCTTGAAGTTGATATTCAGAGAAGATGACCAAGAGTCTTTATTT 1077

OY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTTAT 1054

RESULT 5

US-09-709-126-5/c

; Sequence 5, Application US/09709126

; Patent No. 6319704

; GENERAL INFORMATION:

; APPLICANT: Rubin, Gerald M.

; Pan, Duojia

; Rooke, Jenny

; Yavari, Reza

; Xu, Tian

; TITLE OF INVENTION: KUZ: A No. 6319704el Family of Metalloproteases

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/709,126

; FILING DATE: 08-No. 6319704-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/285,502

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A.

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: B97-081

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2098 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-709-126-5

Query Match 3.4%; Score 36.6; DB 4; Length 2098;

Best Local Similarity 65.1%; Pred. No. 0.87;

Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 816 AGCAACCTCCCTTGGATGATGACAGCAAGATGACAAAGAGTAAAGTTTGA 875

DB 1136 AGCCAAAGCTTGAAGTTGATATTCAGAGAAGATGACCAAGAGTCTTTATTT 1077

DB 1136 AGCCACAAAAGCTTGAAGTTGATATTCAGAGAAGATGACCAAGAGTCTTTATTT 1077

OY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTTAT 1054

RESULT 6

US-09-871-385A-5/c

; Sequence 5, Application US/09871385A

; Patent No. 6399350

; GENERAL INFORMATION:

; APPLICANT: Rubin, Gerald M.

; Pan, Duojia

; Rooke, Jenny

; Yavari, Reza

; Xu, Tian

; TITLE OF INVENTION: KUZ: A No. 6399350el Family of Metalloproteases

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/871,385A

; FILING DATE: 31-May-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/709,126

; FILING DATE: 08-No. 6399350-2000

; APPLICATION NUMBER: 09/285,502

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A.

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: B97-081

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2098 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-871-385A-5

Query Match 3.4%; Score 36.6; DB 4; Length 2098;

Best Local Similarity 65.1%; Pred. No. 0.87;

Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 816 AGCAACCTCCCTTGGATGATGACAGCAAGATGACAAAGAGTAAAGTTTGA 875

DB 1136 AGCCACAAAAGCTTGAAGTTGATATTCAGAGAAGATGACCAAGAGTCTTTATTT 1077

OY 876 AAAATCTTAATAGAAATGCAT 898

DB 1076 AGGCTACTGAAATATTATTTAT 1054

RESULT 7

US-09-737-698B-25

; Sequence 25, Application US/09737698B

Patent No. 6462258
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen
; APPLICANT: Wilkinson, Jack
; TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
; FILE REFERENCE: 38-21(51499)C
; CURRENT APPLICATION NUMBER: US/09/737,698B
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 60/171,173
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 25
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)-(1241)
; OTHER INFORMATION: n= a or g or c or t/u
; OTHER INFORMATION: Act7 promoter polynucleotide sequence and intron
US-09-737-698B-25

Query Match 3.4%; Score 36.4; DB 4; Length 1241;
Best Local Similarity 59.8%; Pred. No. 0.79; Mismatches 41; Indels 0; Gaps 0;
Matches 61; Conservative 0;
QY 361 TAGCGTGTGTTTACTAGAGAGAGCTGTACTTTCTCTATGACTTTAGTGAAGAGTCT 420
DB 1030 TAGGTGTGATGTTTACCAAGTGAATCTAGTGTCTTCTTTGAGATCTGGAAGTTGA 1089
QY 421 TCCATGTTTATGAAAAATTCAGAAATGAGGATGTAGAG 462
DB 1090 ACCTAGTTTCTCAATAATCAACATATGAAGGATGTTTGAG 1131

RESULT 8
US-08-347-340-1
; Sequence 1, Application US/08347340
; Patent No. 5633439
; GENERAL INFORMATION:
; APPLICANT: WALTER, MICHAEL H
; TITLE OF INVENTION: EXPRESSION OF GENES IN TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: DC USA
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,340
; FILING DATE: 06-DEC-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/01098
; FILING DATE: 27-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9211416.4
; FILING DATE: 29-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 9320/212448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/861-3000
; TELEFAX: 202/822-0944

TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6877 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE: Nicotiana tabacum
; ORGANISM: Nicotiana tabacum
US-08-347-340-1
Query Match 3.2%; Score 34.8; DB 1; Length 6877;
Best Local Similarity 44.5%; Pred. No. 4.6; Mismatches 172; Indels 0; Gaps 0;
Matches 138; Conservative 0;
QY 592 AGCTACAAATCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAGAATTTAAT 651
DB 61 ATCAATAAATGTCTAACTGCAATGATCTGTAAATTTGTGTATCAGAGAAATGCTGACT 120
QY 652 GACTCTGTGATGCTTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAA 711
DB 121 GATATATAGTAGGCATTTGAAAGTTGAAAGATGAATTCACACTAAACTCAGATAAGATG 180
QY 712 AAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGCAAGGAGTCGATACATAA 771
DB 181 ATTTTGGGGTAGGCATTTGATTAATGCAAAACCTACTCTCATGAATGTCAGTTTAAC 240
QY 772 ATTCGGATTAATAAAGGAGTGTCTTACACAGCCATATCAGTGGAGCAACCTTCCCTTTG 831
DB 241 AACCAATCAACCAATGAATTTCTAAAAACAAGAAATTTACTGGGAAAAAACAACGTTTCG 300
QY 832 ATCTTTGTAATACAGAGATGACAAAGAGTAAGTTTTTGAAAAATTTCTAAATAGA 891
DB 301 AATAAATAAGAAATATCGTAATTAATTAACCAACAATTTCCATTGGTTATAACTTA 360
QY 892 AATGCATTTG 901
DB 361 TAATCACGTG 370
RESULT 9
US-08-602-036A-1
; Sequence 1, Application US/08602036A
; Patent No. 5789248
; GENERAL INFORMATION:
; APPLICANT: Oeystein, Fodstad
; APPLICANT: Hovig, Eivind
; APPLICANT: Engebraaten, Olav
; APPLICANT: Maelandsmo, Gunhild H.
; APPLICANT: Agrawal, Sudhir
; TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
; TITLE OF INVENTION: METHODS OF INHIBITING METASTATIC CANCER
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States of America
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,036A
; FILING DATE: 16-FEB-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523

```

1 REFERENCE/DOCKET NUMBER:  HYV-039C1P
2
3 TELECOMMUNICATION INFORMATION:
4
5     TELEPHONE: (617) 526-6000
6     TELEFAX: (617) 526-6000
7
8 INFORMATION FOR SEQ ID NO: 1:
9
10    SEQUENCE CHARACTERISTICS:
11
12        LENGTH: 10952 base pairs
13        TYPE: nucleic acid
14        STRANDEDNESS: single
15        TOPOLOGY: linear
16
17    MOLECULE TYPE: genomic DNA
18
19    HYPOTHETICAL: NO
20
21    ANTI-SENSE: NO
22
23    US-08-602-036A-1

```

Query Match	3.2%	Score 34.6;	DB 1;	Length 10952;
Best local similarity	55.4%;	Pred. NO. 6.5;		
Matches 67;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;
QY	410	GTGAAGAGTCTTCCCATGTTTATGAAAAATTCAGAAATAGGGATGTGAGAACCCCAAGC	469	
Db	7700	GTGGAGGAGACCCGCTGTTTCAGCAGCATGAGATATATAGGCATCTCGAATCCACCGT	7759	
QY	470	ACGTGCGCCCTAAAAGGTCATGGATGAGATCAGAGGATTAAAGGAGAAATATCTTTATCGAGA	529	
Db	7760	CTCTGCCATAAAGGTCCTGGATGAGATCTGTTATCAGTCAGAAATCAAGCTCTTTCAGA	7819	
QY	530	G 530		
Db	7820	G 7820		

```

1  RESULT 10
2  US-08-502-374A-1
3  ; Sequence 1, Application US/08502374A
4  ; Patent No. 5872007
5  ;
6  ; GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Fodstad, Oeystein
9  ; APPLICANT: Hovig, Eivind
10 ; APPLICANT: Engebraaten, Olav
11 ; APPLICANT: Maelandsmo, Gunhild H.
12 ; APPLICANT: Agrawal, Sudhir
13 ;
14 ; TITLE OF INVENTION: CAP1-SPECIFIC OLIGONUCLEOTIDES AND
15 ; TITLE OF INVENTION: METHODS OF INHIBITING METASTATIC CANCER
16 ;
17 ; NUMBER OF SEQUENCES: 21
18 ; CORRESPONDENCE ADDRESS:
19 ;

```

```

:      TOPOLOGY: linear
:      MOLECULE TYPE: genomic DNA
:      HYPOTHETICAL: NO
:      ANTI-SENSE: NO
:
US-08-502-374A-1

Query Match          3.2%;      Score 34.6;  DB 2;      Length 10952;
Best Local Similarity 55.4%;      Pred. No. 6.5;
Matches 67;  Conservative 0;  Mismatches 54;  Indels 0;  Gaps

QY  410  GTGAAGAGCTCTTCCCATGTTTATCGAAAAATTCAGAAAATGAGGGATGTAGAGGCCAAGGC 469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    7700  GTGGAGGAGACCCGCTGTGTCAGCAGCATGAGAATATAGGCATTCTCGAATCCACCGT 7759

QY  470  ACGTCCCTAAAGGCTATGGATGAGATCAGGAGGTTTAAGGAGAAATATATCTTTATTCGAGA 529
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    7760  CTCGCCATAAAGGTCCTCGGATGAGATCTGTTATCAGTCAGGAATCAAGCTCTTCAGA 7819

QY          530  G 530
      |
Db    7820  G 7820

```

RESULT 11
US08-642-407A-1
Sequence 1, 5877108
PCT No. 5877108
GENERAL INFORMATION:
APPLICANT: Deystein, Fodstad
INVENTOR: Deystein, Fodstad
APPLICANT: Enggraaten, Olav
INVENTOR: Enggraaten, Olav
APPLICANT: Maelandsen, Gunhild H.
INVENTOR: Maelandsen, Gunhild H.
APPLICANT: Agrawal, Sudhir
INVENTOR: Agrawal, Sudhir
TITLE OF INVENTION: CAPL SPECIFIC OLIGONUCLEOTIDES AND
METHODS OF INHIBITING METASTATIC CANCER
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

Query Match 3.2%; Score 34.6; DB 2; Length 10952;
 Best Local Similarity 55.4%; Pred. No. 6.5;
 Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

NAME/KEY: misc_feature
LOCATION: (1)...(112132)
OTHER INFORMATION: n = A,T,C or G
US-09-741-150-3

Query Match 3.1%; Score 33.8; DB 4; Length 112132;
Best Local Similarity 51.0%; Pred. No. 29;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 265 ATATGCTTTTCTCCCTTCATGTTTCAATAGTCCGCTTCAACAGTAACCAAGCA 324
DB 103641 ATATCCCTTTCTTATATTGAATCCTCACATTCATTTGTCAATATACCAAGTA 103582
QY 325 CGGTACATAGCATGAGTCTTTGTAGGTGTGATCGTGTGTTTACTAGGAAG 384
DB 103581 TCCCGAGAGAGCAGTAGTAAAGATCTTCTTGTGTTTGTGTTTGTGTTTGTGAGACAGA 103522
QY 385 AGCTTGACTTTTCTATGACATTTAGTGAAGAGCTTT 421
DB 103521 GTCCTGCTCTGTAGCTCAGGCTGGAGTGCAGAGGCAT 103485

RESULT 15
US-08-938-546-1/c

Sequence 1, Application US/08938546
Patent No. 6004556

GENERAL INFORMATION:

APPLICANT: Fosberry, Andrew P.
APPLICANT: Lawlor, Elizabeth J.
APPLICANT: Nicholas, Richard O.
TITLE OF INVENTION: No. 6004556el rsbu-1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US

ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,546
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: P50551

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1002 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-938-546-1

Query Match 3.1%; Score 33.4; DB 3; Length 1002;

Best Local Similarity 49.2%; Pred. No. 5;

Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 865 TAAGTTTTTGAATAATCTAAATAGAAATGCAATTTGTCTCGAGCTAAAGCTTCTC 924
DB 865 TAAGTTTTTGAATAATCTAAATAGAAATGCAATTTGTCTCGAGCTAAAGCTTCTC 924

DB 895 TATGTTTATGTTTTTTTAAATATATCTAAAGATTTTTTGTGTTATCTATAAAGTACCTTTCAC 836
QY 925 TTTATTTTCCCTTTTGTAGAATGATTCGGTAGTAGGAATATGGGTGCGAGTATGCACGCTTG 984
DB 835 TATTTCTAGCTTCAGTCACACCATCGTTTAAATGATAATTAATCATCAAGGTATATAG 776
QY 985 GAGTGATTCGCTGTTCTGGGAAGAACAGGAAAGTGTCTCAGCTTTAGAGCAATCT 1043
DB 775 GAATTTCTGTTGTTGATATCGTGTGTTGTAATCTTCTTAACACACTTACCTCTACT 717

RESULT 16

US-09-340-812-1/c

Sequence 1, Application US/09340812

Patent No. 6111078

GENERAL INFORMATION:

APPLICANT: Fosberry, Andrew P.

APPLICANT: Lawlor, Elizabeth J.

APPLICANT: Nicholas, Richard O.

TITLE OF INVENTION: No. 6111078el rsbu-1

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

COUNTRY: US

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/340,812

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/938,546

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O

REGISTRATION NUMBER: 28,354

REFERENCE/DOCKET NUMBER: P50551

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-994-2252

TELEFAX: 215-994-2222

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1002 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-340-812-1

Query Match 3.1%; Score 33.4; DB 3; Length 1002;

Best Local Similarity 49.2%; Pred. No. 5;

Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 865 TAAGTTTTTGAATAATCTAAATAGAAATGCAATTTGTCTCGAGCTAAAGCTTCTC 924
DB 865 TAAGTTTTTGAATAATCTAAATAGAAATGCAATTTGTCTCGAGCTAAAGCTTCTC 924

QY 925 TTTATTTTCCCTTTTGTAGAATGATTCGGTAGTAGGAATATGGGTGCGAGTATGCACGCTTG 984
DB 835 TATTTCTAGCTTCAGTCACACCATCGTTTAAATGATAATTAATCATCAAGGTATATAG 776

QY 985 GAGTGATTCGCTGTTCTGGGAAGAACAGGAAAGTGTCTCAGCTTTAGAGCAATCT 1043
DB 775 GAATTTCTGTTGTTGATATCGTGTGTTGTAATCTTCTTAACACACTTACCTCTACT 717

```
RESULT 17
US-08-938-546-3/c
; Sequence 3, Application US/08938546
; Patent No. 6004556
; GENERAL INFORMATION:
; APPLICANT: Fosberry, Andrew P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Nicholas, Richard O.
; TITLE OF INVENTION: No. 6004556el rsbu-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,546
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/938,546
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1020 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-938-546-3
Query Match 3.1%; Score 33.4; DB 3; Length 1020;
Best Local Similarity 49.2%; Pred. No. 5;
Matches 88; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 865 TAAGTTTGTGAAAATCTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTC 924
DB 913 TAGTTTATGTTTTTAAATATATCTTAAAGTTTTTGTATCTATAAGGTCACCTTCAC 854
QY 925 TTTATTTCTCTTTGTAGATGATTCGGTAGTAGGATATGGGTCGAGTATGCACGTTG 984
DB 853 TATTTCTAGCTTCAGTCACACCCGTTAAATGATATTAATCAATCAAGGTATATAG 794
QY 985 GAGTGATTCGCTGTTCTGGAAGAACAGGAAAGTATGTCACGCTTTAGAGCAATCT 1043
DB 793 GAATTTCTTGTGTGATATCGTGTGTTGTAAGTATCTCTAACACTCTACCTCTAACT 735
; RESULT 18
US-09-340-812-3/c
; Sequence 3, Application US/09340812
; Patent No. 6111078
; GENERAL INFORMATION:
; APPLICANT: Fosberry, Andrew P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Nicholas, Richard O.
; TITLE OF INVENTION: No. 6111078el rsbu-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```

```
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/340,812
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/938,546
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Todd Q
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: P50551
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2252
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1020 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-340-812-3
Query Match 3.1%; Score 33.4; DB 3; Length 1020;
Best Local Similarity 49.2%; Pred. No. 5;
Matches 86; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
QY 865 TAAGTTTGTGAAAATCTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTC 924
DB 913 TAGTTTATGTTTTTAAATATATCTTAAAGTTTTTGTATCTATAAGGTCACCTTCAC 854
QY 925 TTTATTTCTCTTTGTAGATGATTCGGTAGTAGGATATGGGTCGAGTATGCACGTTG 984
DB 853 TATTTCTAGCTTCAGTCACACCCGTTAAATGATATTAATCAATCAAGGTATATAG 794
QY 985 GAGTGATTCGCTGTTCTGGAAGAACAGGAAAGTATGTCACGCTTTAGAGCAATCT 1043
DB 793 GAATTTCTTGTGTGATATCGTGTGTTGTAAGTATCTCTAACACTCTACCTCTAACT 735
; RESULT 19
US-08-938-546-5/c
; Sequence 5, Application US/08938546
; Patent No. 6004556
; GENERAL INFORMATION:
; APPLICANT: Fosberry, Andrew P.
; APPLICANT: Lawlor, Elizabeth J.
; APPLICANT: Nicholas, Richard O.
; TITLE OF INVENTION: No. 6004556el rsbu-1
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```


us-09-438-185a-1_1199590_1200675.rni

Fri Jan 10 12:01:27 2003

```

; Sequence 87, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; TITLE OF INVENTION: Conferring Pest Resistance in Plants
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-0788100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13149 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..13149
; OTHER INFORMATION: /note= "RG2A"
;
; US-09-004-838-87
;
; Query Match 3.0%; Score 33; DB 4; Length 13149;
; Best Local Similarity 48.6%; Pred. No. 20;
; Matches 120; Conservative 0; Mismatches 125; Indels 2; Gaps 1;
;
; QY 818 CAACCTCCCTTGGATCTTTGATAATGACAGAGATGACAAAGAGTAAGTTTGGAAA 877
; DB 7120 CAACATCCATTGGATTGTTGTTGTCACATGGAGTGAATACACACAGTGGTGAAG 7179
; QY 878 AATTCTAAAATAGAAATGCAATTTGTGTCAGAGCTTAAAGCTTCTCTTTATTTTCTTT 937
; DB 7180 AATTATTAAGTGATGACCTTGTGTGATAGCTTTGTGAATCT--CTTTCCACACATCCCATG 7237
; QY 938 TGTAGAAATGATTCGGTCTAGTAGATATGGGTGAGTATGACAGTTGAGTGAATGGCTG 997
; DB 7238 TCTATCTCATCATCTTGAAGAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 7297
; QY 998 TTCTGGAAGAACAGGAGGAAAGATTTCTCAGCTTTTAGAGCAATCTTTCAGATATATCTTT 1057
; DB 7298 TTCAACATTGCTTGGATTGTCTGTTGTCATTTGGGCAAGAGACACAGCATCAGCTTA 7357
; QY 1058 GGTGCA 1064
; DB 7358 AGAACA 7364
;
; US-09-004-838-21
;
; Query Match 3.08; Score 33; DB 4; Length 10815;
; Best Local Similarity 48.6%; Pred. No. 18;
; Matches 120; Conservative 0; Mismatches 125; Indels 2; Gaps 1;
;
; QY 818 CAACCTCCCTTGGATCTTTGATAATGACAGAGATGACAAAGAGTAAGTTTGGAAA 877
; DB 7239 CAACATCCATTGGATTGTTGTTGTCACATGGAGTGAATACACACAGTGGTGAAG 7298
; QY 878 AATTCTAAAATAGAAATGCAATTTGTGTCAGAGCTTAAAGCTTCTCTTTATTTTCTTT 937
; DB 7299 AATTATTAAGTGATGACCTTGTGTGATAGCTTTGTGAATCT--CTTTCCACACATCCCATG 7356
; QY 938 TGTAGAAATGATTCGGTCTAGTAGAATATGGGTGAGTATGACAGTTGAGTGAATGGCTG 997
; DB 7357 TCTATCTCATCATCTTGAAGAGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 7416
; QY 998 TTCTGGAAGAACAGGAGGAAAGATTTCTCAGCTTTTAGAGCAATCTTTCAGATATATCTTT 1057
; DB 7417 TTCAACATTGCTTGGATTGTCTGTTGTCATTTGGGCAAGAGACACAGCATCAGCTTA 7476
; QY 1058 GGTGCA 1064
; DB 7477 AGAACA 7483
;
; RESULT 23
; US-09-004-838-87
;
; GENERAL INFORMATION:
; APPLICANT: Michelmore, Richard W.
; APPLICANT: Shen, Kathy
; APPLICANT: Meyers, Blake
; TITLE OF INVENTION: Procedures and Materials for
; TITLE OF INVENTION: Conferring Pest Resistance in Plants
; NUMBER OF SEQUENCES: 140
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/004,838
; FILING DATE: 09-JAN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/781,734
; FILING DATE: 10-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-0788100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10815 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..10815
; OTHER INFORMATION: /note= "RLG2A"
;
; US-09-004-838-21

```

```

US-08-936-165A-6
: Sequence 6, Application US/08936165A
: Patent No. 6348582
: GENERAL INFORMATION:
: APPLICANT: Black, Michael
: APPLICANT: Burnham, Martin
: APPLICANT: Hodgson, John
: APPLICANT: Knowles, David
: APPLICANT: Lonetto, Michael
: APPLICANT: Nicholas, Richard
: APPLICANT: Pratt, Julie
: APPLICANT: Reichard, Richard
: APPLICANT: Rosenberg, Martin
: APPLICANT: Ward, Judith
: APPLICANT: Weinberg, Martin
: TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
: TITLE OF INVENTION: Polypeptides and Their Uses
: NUMBER OF SEQUENCES: 534
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/936,165A
: FILING DATE: 24-SEP-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/027,032
: FILING DATE: 24-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Glimmi, Edward R
: REGISTRATION NUMBER: 38,891
: REFERENCE/DOCKET NUMBER: P50549
: TELEPHONE: 610-270-4478
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 806 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-936-165A-6
: Query Match 3.0%; Score 32.8; DB 4; Length 806;
: Best Local Similarity 53.6%; Pred. No. 6.7;
: Matches 67; Conservative 0; Mismatches 0; Gaps 0;
: QY 27 CATATTTCTTCTGTAAGGGAANTGATTAAACAGACCTTGGTCGAGCATCCATT 86
: DB 190 GAGACCCCTAATCACTGAAGACTGTTATATATACAGAAATATGATGACTACGTT 249
: QY 87 CTAATTTGAAGTGAATTTACCACTTCATCAAAATGCTTTATGAAAAATAAGTCTCT 146
: DB 250 CATAATTTGAAGAAATTTCTATATCAGAAGGACAAATTTGTAAGCTAAATTTGCTATA 309
: QY 147 GTGGA 151
: DB 310 GATGA 314
: RESULT 25
: US-08-702-080-7/c
: Sequence 7, Application US/08858830
: Patent No. 5965693
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavallie, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/702,080
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 448 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: US-08-702-080-7
: Query Match 3.0%; Score 32.6; DB 1; Length 448;
: Best Local Similarity 60.9%; Pred. No. 5.9;
: Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps
: QY 529 AGTTCGATTACAGAGGCAATGGGAAAGTAACGGTCCTGCACACAGACAAAGTATTGATG 588
: DB 202 AGGCTGCATGACAGACACCATAGGAATCTTGTGCTGCATCGCAGCGCTTGAGG 143
: QY 589 GGGAGCTACAAATCTCGAAACGGGATA 615
: DB 142 GGGTCTTATCCCACTCGACACAGAAA 116
: RESULT 26
: US-08-858-830-7/c
: Sequence 7, Application US/08858830
: Patent No. 5965693
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavallie, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,830

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/702,080

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 448 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cdna

US-08-858-830-7

Query Match 3.0%; Score 32.6; DB 2; Length 448;

Best Local Similarity 60.9%; Pred. No. 5.9;

Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 529 AGTTCGCATTACAGAGCAATGGAAAGTAAAGCTCTGCAACAGAGAGCAAGTATTGATG 588

DB 202 AGCTGTCATGACAGAGACCATAGGAATCTTTGCTGTCATCGCATCGGCTTGAGG 143

QY 589 GGGAGCTCAATCTGAAACGGGATA 615

DB 142 GGGCTTATCCACCTGCACAGGAAA 116

RESULT 27

US-08-858-834-7/c

; Sequence 7, Application US/08858834

; Patent No. 5981222

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John

; APPLICANT: Lavallee, Edward

; APPLICANT: Racie, Lisa

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Spaulding, Vikki

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

; TITLE OF INVENTION: ENCODING THEM

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,834

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/702,080

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 448 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cdna

US-08-858-834-7

Query Match 3.0%; Score 32.6; DB 2; Length 448;

Best Local Similarity 60.9%; Pred. No. 5.9;

Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 529 AGTTCGCATTACAGAGCAATGGAAAGTAAAGCTCTGCAACAGAGAGCAAGTATTGATG 588

DB 202 AGCTGTCATGACAGAGACCATAGGAATCTTTGCTGTCATCGCATCGGCTTGAGG 143

QY 589 GGGAGCTCAATCTGAAACGGGATA 615

DB 142 GGGCTTATCCACCTGCACAGGAAA 116

RESULT 28

US-08-930-285-16

; Sequence 16, Application US/08930285

; Patent No. 6222099

; GENERAL INFORMATION:

; APPLICANT: Regents of the University of Minnesota, et al.

; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ACETYL COA CARBOXYLASE GE

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P. A.

; STREET: P. O. Box 2938

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: PASEQ Version 1.5

; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/930,285

FILING DATE: 13-APR-1998

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/04625

FILING DATE: 04-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Embretson, Janet E.

REGISTRATION NUMBER: 39,665

REFERENCE/DOCKET NUMBER: 600.318US4

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-339-0331

TELEFAX: 612-339-3061

TELEX:

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 638 base pairs

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-930-285-16

Query Match      3.0%; Score 32.6; DB 4; Length 638;
Best Local Similarity 52.6%; Pred. No. 6.9;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 837 TTGATATGACAGACATCAACAGAGCTAGTCTTTTCAAAAATTCATAAATAGAAATGC 896
DB 190 TTGAATATGACATGCTGAGCAGATATTAACAGTTTTTTTATAAAAAACATCATTT 249

QY 897 ATTGTGTCAGCTAAGCTTCCTCTTTTATTTTCTTTTCTAGTAAGATCGGTAGT 956
DB 250 CTAGGAGTTGGACTAAGCTTCTTTCTAGTATGAAGTGCCATGTTTACATGGTCCATTTGT 309

QY 957 AGGAATATGGGTGCG 971
DB 310 GTCAATTTACAGTCG 324

RESULT 29
US-08-484-438-3
; Sequence 3, Application US/08484438
; Patent No. 5811098
; Patent No. 5811098 5780031
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, Gregory D.
; APPLICANT: CLOUSCOU, Jean-Michel
; APPLICANT: SHOYAB, Mohammed
; APPLICANT: STEGALL, Clay B.
; APPLICANT: HELISTR M, Ingeyerd
; APPLICANT: HELISTR M, Karl E.
; TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT PATENT IN RELEASE #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,438
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/323,442
; FILING DATE: 14-OCT-1994
; APPLICATION NUMBER: US 08/150,704
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/981,165
; FILING DATE: 24-NOV-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

; TELEFAX: (212) 869-8664/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5555 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34...3210
; US-08-484-438-3

Query Match      3.0%; Score 32.4; DB 1; Length 5555;
Best Local Similarity 53.1%; Pred. No. 20;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 585 GATGGGAGCTACAATCTCTGAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGA 644
DB 5423 GTTTGGTGGCTAATACGCTGCTAGGGAAACATCTTAAAAAGTGAATTTTGATCAAAATAT 5482

QY 645 ATTTAATCACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCAGAAAGCTTCTTAAGTAAT 704
DB 5483 TTCTTAAGCATATGCTGATAGACTTTCAAACCAAAAAAAGAAAAAAGAAAAA 5542

QY 705 CAAGAAAAAA 714
DB 5543 AAAAAA 5552

RESULT 30
US-08-965-762-15
; Sequence 15, Application US/08965762
; Patent No. 6280963
; GENERAL INFORMATION:
; APPLICANT: KOLTB, Yigal
; APPLICANT: GAVRIAS, Victoria
; TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE
; FILE REFERENCE: 07334/062001
; CURRENT APPLICATION NUMBER: US/08/965,762
; CURRENT FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 3800
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; US-08-965-762-15

Query Match      3.0%; Score 32.2; DB 4; Length 3800;
Best Local Similarity 61.2%; Pred. No. 19;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 856 AACAAAGAGTAAGTTTTTGAATAATCTAAATAGAAATGATGTCGAGCTAAAGC 915
DB 3 AAAGAAGAGTCAGACTTTTAAAGAGTAAACCAAAAAAAGCTTTTTTTAGGACGGATTATACC 62

QY 916 TTGCTTCTCTTTATTTTCTCTTTTGT 940
DB 63 ATAGTCTCTCTTATTTCTCTTTT 87

RESULT 31
US-09-851-896-3/c
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
```



```

1 FILE REFERENCE: RFS-0220
2 CURRENT APPLICATION NUMBER: US/09/851,896
3 CURRENT FILING DATE: 2001-05-08
4 NUMBER OF SEQ ID NOS: 83
5 SEQ ID NO: 3
6 LENGTH: 70000
7 TYPE: DNA
8 ORGANISM: Homo sapiens
9 FEATURE:
10 US-09-851,896-3

```

[illegible]

RESULT 32
 US-09-134-001C-2094
 Sequence 2094, Application US/09134001C
 Patent No. 6380370
 GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND N
 TITLE OF INVENTION: EPIDERMIS FOR D
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134,0
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,96
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,77
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 2094
 LENGTH: 528
 TYPE: DNA
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-2094

[illegible]

RESULT 33
US-08-568-459A-11

```

; Sequence 11, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-Zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Isrigelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-11

```

	Query Match	2.9%;	Score 32;	DB 2;	Length 8220;
	Best Local Similarity	56.7%;	Pred. No. 31;		
	Matches	59;	Conservative	Mismatches	Indels
					Gaps
QY	155	GATGTATTCAAGGGAGACGAATAATCTGATCCAAATCTCTAAAGGGAGAACACGTACGTTA	214		
Db	2789	GGTGTGCGCATGAGAATAGGAACGGAAATGGTCAAAATATTGAAGGAAAAAAACCAACGTCA	2848		
QY	215	TCAATGAAGAGGTGTCCAAGTTCTCTGGAGGAAGCTGATAAGAC	258		
Db	2849	TACAAAAACGTCTTTTACCTCCCGCAGCAGAGAACACATGTGTAC	2892		

RESULT 34
US-08-487-826B-11
; Sequence 11, Application US/08487826B
. patent NO 5993827

GENERAL INFORMATION:
APPLICANT: Sun, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING D
TITLE OF INVENTION: AND PLASM

APPLICANT:	INVENTOR(S):	TITLE OF INVENTION:	BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```
;
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
;
; US-08-487-826B-11
;
; Query Match 2.9%; Score 32; DB 2; Length 8220;
; Best Local Similarity 56.7%; Pred. No. 31;
; Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
;
; QY 155 GATGTATTGAAGGAGAGCAATAATCTGATCCCAATTCCTAAAGGAGAGACAGTACGTTA 214
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2789 GGTGTGCGCATGAGATAGACGGAATGGTCAAAATATTGAAGGAAAAAACAACGTCA 2848
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 215 TCAATGAAGCGTGTCCTGAGGAGAACTGATAAGAAC 258
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2849 TACAAAAACGCTCTTTTACCTCCCGACGAGAACACATGTGTAC 2892
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 35
; US-09-210-288-11
; Sequence 11, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-Zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8220 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
;
; US-09-210-288-11
;
; Query Match 2.9%; Score 32; DB 4; Length 8220;
; Best Local Similarity 56.7%; Pred. No. 31;
; Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
;
; QY 155 GATGTATTGAAGGAGAGCAATAATCTGATCCCAATTCCTAAAGGAGAGACAGTACGTTA 214
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2789 GGTGTGCGCATGAGATAGACGGAATGGTCAAAATATTGAAGGAAAAAACAACGTCA 2848
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; QY 215 TCAATGAAGCGTGTCCTGAGGAGAACTGATAAGAAC 258
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
; DB 2849 TACAAAAACGCTCTTTTACCTCCCGACGAGAACACATGTGTAC 2892
; | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
;
; RESULT 36
; US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-Zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
```

us-09-438-185a-l_1199590_1200675.rni

Fri Jan 10 12:01:27 2003

TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 2.9%; Score 32; DB 2; Length 19124;
Best Local Similarity 56.7%; Pred. No. 44;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 155 GATGTATTGAAGGAGACGAATAATCTCATCAATCTTAAAGGGAGACACGCTAGCTTA 214
DB 10109 GGTGTCCCATGAGATAGGACGGAATGGTCAATATTGAAGGAAAAACAAACGCTCA 10168

QY 215 TCATGAAGCGGTCTCAAGTTCTCTGAGGAGAACTGATAAGAAC 258
DB 10169 TACAAAACGCTCTTTTACCTCCCGACGAGACACATGTGTAC 10212

RESULT 37
US-09-356-952-12/c
Sequence 12, Application US/09356952
Patent No. 6117663
GENERAL INFORMATION:
APPLICANT: Borlack-Sjodin, Ann
APPLICANT: Margarit, S. M.
APPLICANT: Bor-Sogi, Dafna
APPLICANT: Cole, Philip
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
FILE REFERENCE: 600-1-228N
CURRENT APPLICATION NUMBER: US/09/356,952
CURRENT FILING DATE: 1999-07-19
EARLIER APPLICATION NUMBER: 60/093,631
EARLIER FILING DATE: 1998-07-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 43676
TYPE: DNA
ORGANISM: Saccharomyces cerevisiae
US-09-356-952-12

Query Match 2.9%; Score 32; DB 3; Length 43676;
Best Local Similarity 53.1%; Pred. No. 63;
Matches 68; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 771 AATTCGGATTAAGGAGGAGTGTACACAGCCATATCAGTGGAGCAACCTTCCCTTT 830
DB 32382 AATTGAAGTGNAACCACTTTTCAGTTCATCTTCCCAATATTGTGAAGTACTCATCGCGTA 32323

QY 831 GATTCCTTTGATAATGACAGAGATGAACAAAGATAGTTTGTGAAAAATTCATAAATAG 890
DB 32322 AAGGATTCATGCCTTCTTGAATGAAGATATTGACATAGTTTTTGAATAATTTGTAANAAC 32263

QY 891 AATGTCAT 898
DB 32262 TTCTGGAT 32255

RESULT 38
US-08-463-090B-3
Sequence 3, Application US/08463090B
Patent No. 5801015
GENERAL INFORMATION:
APPLICANT: Kondo, Keiji
APPLICANT: Inouye, Masayori
TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
TITLE OF INVENTION: APPLICATIONS
NUMBER OF SEQUENCES: 10

APPLICANT: Cottarel, Guillaume
APPLICANT: Damagnez, Veronique
APPLICANT: Draetta, Guilio
TITLE OF INVENTION: Cell-Cycle Regulatory Proteins from
TITLE OF INVENTION: Human Pathogens, and Uses Related Thereto
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Eliot, LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,090B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIV032.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-7000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1002 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 43..993
US-08-463-090B-3

Query Match 2.9%; Score 31.8; DB 1; Length 1002;
Best Local Similarity 47.3%; Pred. No. 14;
Matches 96; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 87 CTAATTTTGAAGTGAGTTTACCAGTTTCATCAATAATCTTATTGAAATAAAGTCTCTT 146
DB 362 CTAATATGATAAAAGATTTATGAATCAATTTATTCGAGGTATTAACATGTCATCTC 421

QY 147 GTGGAGTTGATGATTTGAAGGAGAGACGATTAATCTGATCCCAATCTTAAGGGAGACAC 206
DB 422 ATCGAGTTTACATCGTGATTTAAACACCAAAATTTATGATTGATAAAGAGGAATT 481

QY 207 GTACGTTATTAATGAAGCGTGTCCAGTTCTCGAGGAGAACTGATGAAGACGGCTCCAT 266
DB 482 TAAATATACAGATTTTGATAGTCGAGCATTTGGAGTTCCCATTAAGACCATATATC 541

QY 267 ATGCTTTTCTTCTTCATGGTTT 289
DB 542 ATGAAGTTGTCACCTTTATGGTAT 564

RESULT 39
US-07-667-276A-1/c
Sequence 1, Application US/07667276A
Patent No. 5470971
GENERAL INFORMATION:
APPLICANT: Kondo, Keiji
APPLICANT: Inouye, Masayori
TITLE OF INVENTION: STRESS-INDUCED PROTEINS, GENES CODING
TITLE OF INVENTION: THEREFOR, TRANSFORMED CELLS OF ORGANISMS, METHODS AND
TITLE OF INVENTION: APPLICATIONS
NUMBER OF SEQUENCES: 10

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/567,276A
; FILING DATE: 11-MAR-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 377,5351P
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1584 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: S288C
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 475..1104
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1
; OTHER INFORMATION: /note= "Base #1 of Sequence No. 5470971.1
; OTHER INFORMATION: corresponds to base -474 of the sequence listed in
; OTHER INFORMATION: Figure 4 of the application"
US-07-667-276A-1

Query Match 2.9%; Score 31.8; DB 1; Length 1584;
Best Local Similarity 51.9%; Pred. No. 17;
Matches 95; Conservative 0; Mismatches 87; Indels 1; Gaps 1;

QY 606 AAACGGGATAATAGGAAATCTGTTCTAATTCGAAGAAATTTAATGACTGCTTGTGATGCT 665
DB 1296 AATTGTCAGAAGGAACATTTCTAGGATGGGAGAGACGCTAATCTCGAAAAAGA 1237
QY 666 CTAGAANTCAGAAGGAGACCGAAGCTTCTAAGTAATCAAGAAAAATCTTGACAAAT 725
DB 1236 GAACGTTTTTAGAAGAGTTTTAGAGTTGATATATTAATGAAGAAAAATTTTCGA-AC 1178
QY 726 AGCTTTTCCATAGAGAAACCGTAGGAAAGAGGAGTCGATACCAATAAATCCGATTAAAAA 785
DB 1177 AACAAATFACATAAGAGAGGTATAATAATGAGGTAATTTTTCAAAATATATTGTTGTAA 1118
QY 786 AGG 788
DB 1117 AGG 1115

RESULT 40
US-08-642-274D-51
; Sequence 51, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:

; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
; FILE REFERENCE: 229000033
; CURRENT APPLICATION NUMBER: US/08/642,274D
; CURRENT FILING DATE: 1996-05-03
; NUMBER OF SEQ ID NOS: 220
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: genomic
US-08-642-274D-51

Query Match 2.9%; Score 31.6; DB 4; Length 679;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 79; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 866 AAGTTTTTGAAAAATCTAAATAGAAATGCAATTTGTGTCGAGCTAAAGCTTGTCTCTCT 925
DB 83 AAGTTTGCATAGTTCATATAATTTAGCTAGCTTTTATATATATATATAAGTAAATTTAG 142
QY 926 TTATTTTCCTTTTGTAGAAATGATCGGTAGTAGGAATATGGGTGCGAGTATGCACGTGG 985
DB 143 TGTATTACCTTAATTTGAGTGATCTTTTAGATGATTTAGTATTTGTAAATATAATTAA 202
QY 986 AGTGATGGCTGCTTCTGGAAGACAGGGAAGAGTGATG 1023
DB 203 ATTGGTTGCTTTTCTTGAGGAGTAGAAGTTGCTGG 240

RESULT 41
US-08-952-014C-51
; Sequence 51, Application US/08952014C
; Patent No. 6265158
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; TITLE OF INVENTION: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 6265158thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,014C
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290,00028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 679 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-952-014C-51
```

```

Query Match      2.9%; Score 31.6; DB 4; Length 679;
Best Local Similarity 50.0%; Pred. No. 13; Indels 0; Gaps 0;
Matches 79; Conservative 0; Mismatches 79;

      Oy      866  AGCTTTTGAATAATTCCTAAATAGAAATGCAATTTGTCGACGCTAAACCTTGCTTCTCT 925
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB      83  AGTTTGCAATAGTTTCATATAAATTAGCTAGCTTTATATATGTATATAGTTAAATTTAG 142

      Oy      926  TTATTTTCTCTTTCTAGCAATGATTCGGTAGTAGGAATATGGGTCGAGTAGTCACGTTGG 985
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB      143  TGTATTTACCTTAAATTTGAGTGATCTCTTTAGATGTATTTAGTATTTGTAAATATAATTAA 202
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

      Oy      986  AGTGATTGGCTGTTCTTGGAAGACAGGGAAGTGATTG 1023
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      DB      203  ATTGGTTGTGTTTCTTGAAGGAGTAGAAGTTGCTG 240
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 42
US-08-858-052-2
; Sequence 2, Application US/08858052
; Patent No. 5849498
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Guegler, Karl J.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Nell C.
; TITLE OF INVENTION: HUMAN 3-HYDROXYISOBUTYRUL-COENZYME A
; TITLE OF INVENTION: HYDROLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,052
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0294 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1734 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: U937NOT01
; CLONE: 1187
; US-08-858-052-2

Query Match      2.9%; Score 31.6; DB 2; Length 1734;
Best Local Similarity 54.2%; Pred. No. 20; Indels 0; Gaps 0;
Matches 64; Conservative 0; Mismatches 54;

Oy      143  TCTTGTGGAGTTGATGATTGAAGGAAGCAAGTAATCTGTATCCAAATTCCTAAAGGAGA 202

```

```

Db      956  TCITTTGCCCTAGAGCAATTAAGGTAATTAATAAAATCTCTCCAACATCTCTAAAGATC 1015
QY      203  ACAGCTACGTTATCATGAAGCGTGTCCAAAGTTCCTCGAGGAGAACTGATGAAGAAGCG 260
Db      1016  ACATAAGGCAACTCATGAGGGGTCTTCAAGACCTTCCAAAGNAGTAACTATGG 1073

RESULT 43
US-09-200-284-2
; Sequence 2, Application US/09200284
; Patent No. 6110507
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Guegler, Karl J.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN 3-HYDROXYISOBUTYRUL-COENZYME A
; TITLE OF INVENTION: HYDROLASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/200.284
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/858,052
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/POCKET NUMBER: PF-0294 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1734 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: U93/NOT01
; CLONE: 1187
US-09-200-284-2

Query Match 2.9% Score 31.6; DB 3; Length 1734;
Best Local Similarity 54.2%; Pred. No. 20;
Matches 64; Conservative 0; Mismatches 54; Indels 0; Gaps

QY      143  TCITGTGGAGTTGATGTATTCAGGGAAGCAAGTAATCTGATCCAATCTTAAGAGGAGA 202
Db      956  TCITTTGCCCTAGACCAATTAAGGTAATTAATAAAATCTCTCCAACATCTCTAAAGATC 260
QY      203  ACAGCTACGTTATCATGAAGCGTGTCCAAAGTTCCTCGAGGAGAACTGATGAAGAAGCG 260
Db      1016  ACATAAGGCAACTCATGAGGGGTCTTCAAGACCTTCCAAAGNAGTAACTATGG 1073

RESULT 44
US-09-484-970B-60

```

; Sequence 60, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO: 60
; LENGTH: 13977
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 229357.11CBI
; NAME/KEY: unsure
; LOCATION: 11721-11761, 12294, 13969
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-60

Query Match 2.9%; Score 31.6; DB 4; Length 13977;
Best Local Similarity 46.1%; Pred. No. 50;
Matches 106; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 765 ACCATAAATCCGATTAAAGAGGAGTGTACACAGCCATATCAGTGGAGCAACCTT 824
DB 11867 ACACCAATGACTACTACACACTGCTCATTTGAGAACTGTGAGTGGGGAGGCTT 11926
QY 825 CCCTTTCATCTTTGTAATGACAGAAAGAGTAAAGTTTGGAAATTTCTA 884
DB 11927 GAGTTTCATTCATATCTATATGCTATAGTATATAAATCTATATAGAT 11986
QY 885 AATAGAAATGCAATTTGTCGAGCTAAAGCTTCTCTCTTTATTTTCTTTGTAGAA 944
DB 11987 ANAGATAGCAATTTCTATAGACGACTTTTTCATTTTAAATGTTTCATGTCACATC 12046
QY 945 TGATCGGTAGTAGGAATATGGGTGCGAGTAGTACGACCTTGGAGTGATGG 994
DB 12047 CTAATAGAAAGAAATTAATCTTCTAGTACGTCATCCAGGCTTACCTGCTTGG 12096

RESULT 45
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345, 2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)

; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1
Query Match 2.9%; Score 31.6; DB 4; Length 168575;
Best Local Similarity 50.0%; Pred. No. 1,4e+02;
Matches 105; Conservative 0; Mismatches 104; Indels 1; Gaps 1;
QY 684 GACCGAAGCTTCTAAGTAATCAAGAAAAAATCTTGGACAAATAGCTTTTCCATAGAGAAA 743
DB 78118 GAAGAAGCCAGATTTTATAAGTAAATAATCTGCTTGTGCTTTTCCA-AATTAGA 78176
QY 744 ACCGTGAGGAAAGGAGTCGATACCATATAATTCGGATTAAAAAAGGAGTGTTCACACAG 803
DB 78177 AGAGAAATGTATCATCTTAATACAGCAGATTGAGTTATATAAAGACCTACTCCATCCAA 78236
QY 804 CCATATCAGTGGAGCAACCTTCCCTTTTGATTTCTTTGTAATGACAGAGAGATGAACAAAGA 863
DB 78237 AAAATTGAGTGAATAAAAAAGAAATTCGACTTACTTGTAAAGAGAAAAAGATTGCCAAGGC 78296
QY 864 GTAAGTTTTTGAATAATTCATAAATAGAAA 893
DB 78297 TTGCAGACTTCTGAGGTGTTAAATAACAA 78326

RESULT 46
US-09-428-034-1/c
; Sequence 1, Application US/09428034
; Patent No. 6428996
; GENERAL INFORMATION:
; APPLICANT: Cheng, Kuo-Joan
; APPLICANT: Lio, Jin-Hao
; APPLICANT: Tsai, Cheng-Fang
; APPLICANT: Hsu, Yih-Chin
; TITLE OF INVENTION: CELLULOSE ENZYMES
; FILE REFERENCE: 08919/036001
; CURRENT APPLICATION NUMBER: US/09/428,034
; CURRENT FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1364
; TYPE: DNA
; ORGANISM: Piromyces rhizinflata
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1128)
US-09-428-034-1

Query Match 2.9%; Score 31.4; DB 4; Length 1364;
Best Local Similarity 52.7%; Pred. No. 21;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 829 TTGATCTTTGATTAATGACAGAAAGATGAACAAAGAGTAAGTTTGGAAAAATCTTAAAT 888
DB 1241 TTTTGTGATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1182
QY 889 AGAAATGCAATTTGTGCGAGCTAAAGCTTCTCTCTTTATTTTCTTTTGTAGAATGAT 948
DB 1181 ATTATATTTTATATATATGTAATAATAATTTTTTTTTTTTAAATTTATTAGAAATGAT 1122
QY 949 TCGGTAGTA 957
DB 1121 GGATTAGCA 1113

RESULT 47
US-09-004-838-116
; Sequence 116, Application US/09004838
; Patent No. 6350933
; GENERAL INFORMATION:

Search completed: January 9, 2003, 11:51:00
Job time : 476 secs

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match          2.9%; Score 31.2; DB 4; Length 289;
Best Local Similarity 9.3%; Pred. No. 12;
Matches 26; Conservative 111; Mismatches 142; Indels 1; Gaps 1;

QY 508 GGAGAAATTCCTTTATCGAGATTCGCATTACAGAGCGCAATGGGAAGTAACGGTCCTG 567
DB 3 RGRGRACRARRURURARCRURURURURARCRURURURARCRURURARCRURURARCR 62
QY 568 CAACAGAGACAAGTATTGATGGGGAGCTACAAATCCGAAACGGGATATAGGAAAATG 627
DB 63 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 122
QY 628 TGTTCCTAAATTCGAAGAATTTAATGACTGCTGATGCTTAAGAAATCAGAAAGAGACC 687
DB 123 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 182
QY 698 GAAAGCTTCTAAGTAATCAAGAA-AAAATCTTGGACAATAGCTTTTCCATAGAGAAAACC 746
DB 183 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 242
QY 747 GTAGGAAAGGAGTCGATACCATATAATTCGGATTAAAAAA 786
DB 243 RURCRURURGRGRCRUAIAAAAAAAAAAAAAAAAAAAAAA 282

RESULT 50
US-09-134-001C-1809
; Sequence 1809, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1809
; LENGTH: 999
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1809

Query Match          2.9%; Score 31.2; DB 4; Length 999;
Best Local Similarity 53.2%; Pred. No. 21;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 307 AACAGTAACCAACCAACGCGTACATAGCATCATGCTTTGTAGGTTTGTATCGG 366
DB 373 AAAGCTGACCAATATAGCAAAATATTCTCTGTTAATACAGAAATATTGTATGATGAC 432
QY 367 TTGTTTTTTTACTAGGAAGAGCTTGTACTTTTCTATGACTTTTATGAGAGCTTCCCAT 426
DB 433 ATGGTTTTTTTGACAGTAAGCGCGATTATTTTCTACTGATTTTAGAGGGTATTCTACACAA 492
QY 427 GTTT 430
DB 493 CCTT 496
```


Fri Jan 10 12:01:29 2003

us-09-438-185a-l_1199590_1200675.rst

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 06:35:28 ; Search time 1940 Seconds
(without alignments)
9066.133 Million cell updates/sec

Title: US-09-438-185a-l_COPY_1199590_1200675

Perfect score: 1086

Sequence: 1 ttggcaagatcctcaaac.....gtctctgcagtagtgac 1086

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estnu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_estc.*
- 9: gb_estl.*
- 10: gb_est2.*
- 11: gb_est3.*
- 12: gb_est4.*
- 13: gb_est5.*
- 14: gb_estfun.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss.*
- 19: em_gss_inv.*
- 20: em_gss_pln.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_oth.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	61	5.6	1101	CNS0039G	AL063921 Drosophil
C 2	51.4	4.7	997	CNS005TE	AL060767 Drosophil
C 3	51.2	4.7	1101	CNS0182P	AL108811 Drosophil
C 4	50.2	4.6	1101	CNS017KX	AL108171 Drosophil
C 5	50	4.6	987	CNS017LB	AL108185 Drosophil
C 6	48.4	4.5	1101	CNS0106X	AL098595 Drosophil

C	7	48.2	4.4	1001	CNS0155H	AL105023 Drosophil
	8	48	4.4	1204	CNS016E2	AL106628 Drosophil
	9	47.2	4.3	615	BJ345426	BJ345426 BJ345426
	10	47.2	4.3	635	BJ345414	BJ345414 BJ345414
	11	46.6	4.3	964	CNS006N9	AL065781 Drosophil
	12	45.6	4.2	588	BJ135240	BJ135240 BJ135240
	13	45.2	4.2	1101	CNS0100X	AL098379 Drosophil
	14	45	4.1	348	BJ340360	BJ340360 BJ340360
	15	45	4.1	485	BJ340295	BJ340295 BJ340295
C	16	45	4.1	529	AU038515	AU038515 AU038515
	17	45	4.1	546	BJ400438	BJ400438 BJ400438
	18	45	4.1	547	BJ402020	BJ402020 BJ402020
	19	45	4.1	562	BJ376869	BJ376869 BJ376869
	20	45	4.1	584	BJ341943	BJ341943 BJ341943
	21	45	4.1	590	BJ398603	BJ398603 BJ398603
	22	45	4.1	596	BJ376868	BJ376868 BJ376868
	23	45	4.1	597	BJ401886	BJ401886 BJ401886
	24	45	4.1	598	BJ402538	BJ402538 BJ402538
C	25	45	4.1	654	C93147	C93147 C93147 Dict
	26	45	4.1	654	BJ343976	BJ343976 BJ343976
	27	45	4.1	663	BJ375941	BJ375941 BJ375941
	28	45	4.1	664	BJ401060	BJ401060 BJ401060
	29	45	4.1	665	BJ434901	BJ434901 BJ434901
	30	45	4.1	675	BJ401293	BJ401293 BJ401293
	31	45	4.1	680	BJ343093	BJ343093 BJ343093
	32	45	4.1	688	BJ399159	BJ399159 BJ399159
	33	45	4.1	691	BJ341290	BJ341290 BJ341290
	34	45	4.1	698	BJ401883	BJ401883 BJ401883
	35	45	4.1	718	BJ400818	BJ400818 BJ400818
	36	45	4.1	722	BJ401195	BJ401195 BJ401195
	37	45	4.1	724	BJ341261	BJ341261 BJ341261
	38	45	4.1	734	BJ343890	BJ343890 BJ343890
	39	45	4.1	742	BJ399243	BJ399243 BJ399243
	40	45	4.1	744	BJ347280	BJ347280 BJ347280
	41	45	4.1	760	BJ401342	BJ401342 BJ401342
	42	45	4.1	765	BJ371948	BJ371948 BJ371948
	43	44.6	4.1	988	AL066743	AL066743 Drosophil
C	44	44.2	4.1	958	CNS0072R	AL066801 Drosophil
	45	44	4.1	1101	CNS0074D	AL106896 Drosophil
C	46	43.8	4.0	896	CNS016L1	A2550349 ENTM79TF
C	47	43.8	4.0	899	AZ527921	A2527921 ENTM79TF
C	48	43.8	4.0	914	AZ6687856	AZ6687856 ENTM72TF
C	49	43.8	4.0	924	AZ690917	AZ690917 ENTM65TF
C	50	43.8	4.0	944	AZ549336	AZ549336 ENTED22TF
C	51	43.8	4.0	1101	CNS014MG	AL104338 Drosophil
	52	43.4	4.0	484	BJ344899	BJ344899 BJ344899
C	53	43.4	4.0	729	C93903	C93903 C93903 Dict
C	54	43.4	4.0	1101	CNS00FMC	AL070972 Drosophil
C	55	43.2	4.0	883	BM416099	BM416099 OP21184 M
C	56	43.2	4.0	1101	CNS00B3U	AL056389 Drosophil
C	57	42.8	3.9	901	CNS00KML	AL078168 Drosophil
C	58	42.8	3.9	1101	CNS00ESI	AL069797 Drosophil
C	59	42.6	3.9	791	AL523778	AL523778 AL523778
C	60	42.4	3.9	859	CNS004YY	AL055406 Drosophil
C	61	42.2	3.9	952	CNS006V8	AL065967 Drosophil
C	62	42.2	3.9	1067	CNS00C08	AL058564 Drosophil
C	63	42.2	3.9	1101	CNS016T6	AL107172 Drosophil
	64	42	3.9	757	AL566489	AL566489 AL566489
	65	42	3.9	877	AL514453	AL514453 AL514453
C	66	42	3.9	1101	CNS00FVE	AL071298 Drosophil
C	67	41.8	3.8	886	CNS0075DF	AL29961 clone BA0
C	68	41.6	3.8	1001	CNS0064G	AL062781 Drosophil
C	69	41.6	3.8	1025	CNS014J2	AL104216 Drosophil
C	70	41.6	3.8	1101	CNS0181N	AL108773 Drosophil
C	71	41.4	3.8	567	BE848353	BE848353 w37f03 Y
C	72	41.4	3.8	628	BB631807	BB631807 BB631807
C	73	41.4	3.8	1101	CNS00EEF1	AQ050533 nbdx0004D
C	74	41.2	3.8	801	AQ050553	AQ050553 nbdx0004D
C	75	41.2	3.8	905	CNS00D59	BI600924 603249264
C	76	41.2	3.8	1067	BI600924	BI600924 603249264
C	77	41	3.8	748	BJ141956	BJ141956 BJ141956
C	78	41	3.8	757	AL514421	AL514421 AL514421
C	79	41	3.8	772	BH596443	BH596443 BOGCK93TF

[illegible]

/clone="BACN37D10"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

BASE COUNT 274 a 268 c 128 g 73 t 358 others

ORIGIN

Query Match 4.7%; Score 51.2; DB 17; Length 1101;
Best Local Similarity 20.4%; Pred. No. 0.022;
Matches 118; Conservative 209; Mismatches 249; Indels 2; Gaps 1;

Qy 434 AAAAATTCAGAAATGAGGATGTAGAGCCAGGACGCTGCGCTAAAAGTCATGGATG 493
Db 525 AAAAATTCAGAAATGAGGATGTAGAGCCAGGACGCTGCGCTAAAAGTCATGGATG 584
Qy 494 AGATCAGGAGTAAAGAGAAATATCTTTATTCGAGATTCGCGATTCACAGAGCAATGGGA 553
Db 585 DRWDARGAARAAADAAARAAKATWAAARAAARAAARAAARAAARAAARAAARAAAR 644
Qy 554 AAGTAAGCTCTGCAACAGAGACAAGTATTTGATGGGAGCTACAAATCCTGAAACGGGA 613
Db 645 AAAAGGGWGTGKGAADGGGAAADAWRRGRGRRAARAAARAAARAAARAAARAAAR 704
Qy 614 TAATAGGAAATATGCTTTCTTAATTCGAGAAATTTAATGACTGCTGATGCTTAAGAA 673
Db 705 AAAAAAADAAGKRRKKKGGKDGKATKTAAWAKGRKDWGTATAWTTWATWATW 764
Qy 674 TCAGAAAGGAGACGCAAGCTCTTAAGTAATCAAGAAATATCTTCGCAATAGCTTTTC 733
Db 765 AATDADAKRAAGRRKRDARKTARDGGRARTRRRAAAGGKRRAGARARAAARADD 824
Qy 734 CATAGACAAACCGTGCAGAAAGAGTCGATACCATTAATTCGATTAATAAAGGAGTG 793
Db 825 RMDAWAARAAARAAARAAATWTRDGMWDDWDDWTRDWDWDDWTTAAWDDARARARRRR 884
Qy 794 CTTACACAGCCATATCAGTGAGCAACCTTCCCTTTGATTTGATTTGATTAATGACAGA 853
Db 885 RRRRRRAARAAADDTDFKRW--ADATTDKDTTKTWTWTTDDDDWDAKDRDRAAKADG 942
Qy 854 TGAACAAGAGTAAGTTTGTGAAATTTCTAATAATAGAAATGATTTGTCGAGCTAAA 913
Db 943 AAWWRDARDWAATAKDDGKWKWGRGRGKGRGKGRGKGRGKGRGKGRGKGRGKGRGK 1002
Qy 914 GCTTGCTCTCTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 973
Db 1003 WWWWTRTKWDDWDDGRCGRGWTRKRWGAWRADAWRDDTDGDKDTRADKDKRTDTT 1062
Qy 974 TATGCAAGTTGGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1011
Db 1063 KRGGDDWRDRDKKRRDKKGGDKTKDATTWDDDDARD 1100

RESULT 4
CNS017KX/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence sp6 end of BAC
BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -

collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11

FEATURES
Location/Qualifiers
Source 1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37F10"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

BASE COUNT 247 a 84 c 159 g 263 t 348 others

ORIGIN

Query Match 4.6%; Score 50.2; DB 17; Length 1101;
Best Local Similarity 22.4%; Pred. No. 0.039;
Matches 91; Conservative 135; Mismatches 181; Indels 0; Gaps 0;

Qy 380 GGAAGAGCTGTACTTTTCTATGACTTTTAGTGAAGAGCTTCCCATGTTTATGAAAAAT 439
Db 1048 GGRKADTKRWRGAATRRARWATAATTATADTTWKTGTRTKRWKWWARAAKWRD 989
Qy 440 TCAGAAATGAGGATCTAGACCCAGGACGCTCCCTAAAGTCATGATGATGATCA 499
Db 988 WWRDTRDWDKWDTKKDDRWTKWATRGDGRAGRAKRDRDRGRRRRRRTRKRW 929
Qy 500 GGAGTTAAGGAGAAATATCTTTATCGAGATTCCTCATGAGGCAATGGGAAAGTAA 559
Db 928 RRAAWTATWTTWGGTTKGAWTKTKRKGGAADWDRDMDKRAAAAGKDDG 869
Qy 560 CGGCTCGCAACAGACAGCAAGTATGATGGGAGCTACAAATCCTGAAACGGGATATAG 619
Db 868 KGKARTWGAAGAAAAAWDTKATATKRRKRTXGRAWWAAGTRTTWDAATATKRAAAA 809
Qy 620 GAAATATGTTTCTTAATTCGAAGAAATTAATGACTGCTTGTGCTTAAGAATCAGAA 679
Db 808 AGRRRAAAWAKDRCTRGKRGKRGTDGKGTWKTRTTWTKTAAARRARAAAKGWT 749
Qy 680 AGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGACAAATAGCTTTTCCATAGA 739
Db 748 TKTGRRARAADWATTTTRKGAATDRGAAAAAATTRGAAAAATGKTTDAGAAAAATA 689
Qy 740 GAAACCGCTGAGGAGGAGTCGATACCAATAATTCGATTAATAAAA 786
Db 688 AWAWRRTWTDWAAARAAAGAAWAKAAWADAADAKKGGRRAKRA 642

RESULT 5
CNS017LB
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence sp6 end of BAC
BACN37A07 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -

<http://www.edgp.ebi.ac.uk>. This *Drosophila* melanogaster BAC library (Dros BAC) was made by Alain Billaud at CPHF (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBac11.

FEATURES	source	location/Qualifiers
		1. 987
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone="BACN37A07"
		/clone_lib="DrosBAC"
		/plasmid="pBelobAC11"
		/note="end : Sp6"
BASE COUNT	224 a	122 c 180 g
ORIGIN	211 t	250 others

RESULT 6
CNSO106X
Locus
DEFINITION
CNSO106X linear GSS 26-JUL-1999 Rosophilla melanogaster genome survey sequence T7 end of BAC BACN03K20 of DrosBAC library from Rosophilla melanogaster (fruit fly); genomic survey sequence.
Accession
Version
Keywords
Source
Organism
Rosophilla melanogaster. <i>Drosophila melanogaster</i> . Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphydroidea; Drosophilidae; Rosophilla. 1 (bases 1 to 1101) GSS.
Reference
Authors
Title
Journal
Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
Comment
- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Rosophilla Genome Project (EDGP - http://www.edgp.ebi.ac.uk -. This Rosophilla melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Pavan. It has been constructed in the vector

[illegible]

CNS0155H DNA 1001 bp linear GSS 26-JUL-1999
 CNS0155H Drosophila melanogaster genome survey sequence sp6 end of BAC
 LOCUS BACN13023 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 DEFINITION
 ACCESSION AL105023
 VERSION AL105023.1 GI:5617037
 SOURCE GSS.
 ORGANISM Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha.
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1001)
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr/)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgproject.org.uk/> - This Drosophila melanogaster BAC
 collaboration was made by Alain Billaut at CEPH (Centre

d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

Location/Qualifiers
 source 1..1001
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN13C23"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : SP6"

BASE COUNT 266 a 219 c 134 g 150 t 232 others
 ORIGIN

Query Match 4.4%; Score 48.2; DB 17; Length 1001;

Best Local Similarity 28.5%; Pred. No. 0.13; Indels 0; Gaps 0;
 Matches 145; Conservative 106; Mismatches 257;

Qy 380 GGAAGAGCTTGACTTTTCTATGACTTTTATGAGAGTCTTCCCATCTTTATGAAAAAT 439

Db 494 SGRGCGGGGTDTYKKBKBBYKKGMBGADARAARKTGAGTAAAWRAAGACAAA 553

Qy 440 TCAGAAAATGAGGATGTAGAACCAAGCGCGCTTAAAGGTCTATGATGAGATCA 499

Db 554 AGRRVRRWGDAGAGAGATAGAGAAAAGAAAGRGMAAABDAGRASAAARKTKWG 613

Qy 500 GGAGTTAAGGAGAAATTTATCTTTCGAGAGTTCGCATTACAGAGCAATGGGAAGTAA 559

Db 614. ARGAAAAAGWDADAATWNRKAAARAADADWAAAAACAANAANVHNAWAARAKAA 673

Qy 560 CGTCTCGCAACAGACAGCAAGTATGTGGGAGCTTACAATCTCGAAGCGGATATAG 619

Db 674 ARRAAAAAAGADGAAAAAAGRAAARDARAANAATAAANAATAAARAARADAR 733

Qy 620 GAAAAATGTTCTTAATTCGAGAAATTTAATGACTGCTTGATGCTTAAGAAATCAGAA 679

Db 734 AAAAAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 793

Qy 680 AGGACCGGAAAGCTTCTAAGTAATCAAGAAAAATCTTGGCAATAGCTTTCCATAGA 739

Db 794 AATAATTTAAWAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 853

Qy 740 GAAACCGTCGAGGAAGGTCGATACCATATAAATTCGGATTAAAAAGGGAGTGTACA 799

Db 854 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 913

Qy 800 ACAGCATATCAGTCGAGCAACCTTCCCTTTGATTTCTTGAATAATGACAGAGATGA 859

Db 914 WAWADTTTWTWTWAAWAAATAATWTATTAATTAATAATAATAATAATAATAATW 973

Qy 860 AAGAGTAAGCTTTTCAAAAATCTTAAA 887

Db 974 AWAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1001

RESULT 8

CNS016E2/c 1204 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
 DEFINITION BACN15A12 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106628

VERSION AL106628.1 GI:5622852

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

TITLE Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1204)

Genoscope.

Direct Submission

JOURNAL

Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

Location/Qualifiers
 source 1..1204
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN15A12"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /note="end : T7"

BASE COUNT 298 a 172 c 106 g 316 t 312 others
 ORIGIN

Query Match 4.4%; Score 48; DB 17; Length 1204;

Best Local Similarity 18.2%; Pred. No. 0.14; Indels 0; Gaps 0;
 Matches 66; Conservative 146; Mismatches 150;

Qy 699 AGTAATCAAGAAAAATCTTGGACANTAGCTTTTCCATAGAGAAACCTGAGGAAGGA 758

Db 1189 AKWAKWAKATAKAWKAKAKAKAAAAAADAADAAAKKAAKAAKAAKAAKAAK 1130

Qy 759 GTCGATACCATTAATTCGATTAAAAAGGAGTGTTCACACAGCCATATCAGTGGAGC 818

Db 1129 AKAADDAADAAAAAKAKAKAAAAAADAADAAKAAKAAKAAKAAKAAKAAK 1070

Qy 819 AACCTTCCCTTTGATTTTGTATATACAGAGAGTGAACAAGAGTAAGTTTGTAAAA 878

Db 1069 DDKDKAAADAKKKKKAKADADKAADKDKKAKADAWDDADDAKAKKKKKKKKK 1010

Qy 879 ATTCTAAAATGAAATGCAATTTGTGTCGAGCTTAAAGTCTTCTTCTTATTTCTCTTT 938

Db 1009 ATDADWAAAKKKKKKAWKKKKKKAKAKAKKKKKKKKKKKKKKKKKKKKK 950

Qy 939 GTAGAATGATTCGGTAGTAGGAATATGGGTTCGAGTATGCACCTTGGAGTGTGCTG 998

Db 949 AAADAAAAAANKKKKKKDAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 890

Qy 999 TCTGGAGAACAGGAAAGTGAATGCTCTCAGCTTTAGAGCAATCTTTCAGATATATTG 1058

Db 889 AAKKKDKKDAADKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 830

Qy 1059 GG 1060

Db 829 GG 828

RESULT 9

BJ345426

LOCUS BJ345426

DEFINITION BJ345426 Dictyostellium discoideum cDNA library, AF Dictyostellium

ACCESSION BJ345426

VERSION BJ345426.1 GI:19215933

KEYWORDS EST.

SOURCE Dictyostellium discoideum.

ORGANISM Dictyostellium discoideum.

REFERENCE 1 (bases 1 to 615)

AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.

TITLE Urushinara, H., Tanaka, Y., Kohara, Y., and Shin-i, T.

Full length cDNA of Dictyostellium discoideum at the aggregation

stage

Unpublished (2002)

Genoscope.

Contact: Tadasu Shin-i

Center For Genetic Resource Information

Query Match 4.3%; Score 47.2; DB 13; Length 635;
Best Local Similarity 59.8%; Pred. No. 0.22;
Matches 79; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

DB 51 / AAAAAAATAAAATTTGGGAGTCTTGTGACCAATAGCTTTTCNTAGAGAAAACCGTGAGGAAGAAGACT 760

Qy 701 TAATCAAGAAAAATCTTGGACAATAGCTTTTCNTAGAGAAAACCGTGAGGAAGAAGACT 760

Dh 577 TAAAAAAAAAAAAAADGGTGTGTAATTGTTTAKRTWMDAAAAAAATATATAAAAAAW 636

59

Matches 79; conservative 9; attendances

```

Qy 761 CGATACATAAATCCGATTAAAAAGGAGTCTTACACAGCCATATCAGTGGAGCA 820
    ||| | ||| : : : ||| : ||| : : ||| : : ||| : : ||| : : |||
Db 637 AAATAAAAAAANDKAKWAATAATATAAAATAGTGTAGTAWTAKTAAAAATATAAAAA 696
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 821 CCTTCCCTGTTGTTCTTGTATTAATACAGAGATGACAAAGAGTAGTGTGTTTGA 880
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 697 ADAADAADAATAAKTKWKAADAADAAGAGGAKKDAADAAAAAAGAGAGAKGTATADA 756
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 881 TCTAAAAATGAATACATTTGTGTCGA 907
    : ||| ||| : : : : : : : : : : : : : : : : : : : : : :
Db 757 DAAAAADTAKAKWDNADAKDDDKRAGA 783

RESULT 12
LOCUS BJI135240 588 bp mRNA linear EST 23-JAN-2002
DEFINITION BJI135240 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk1097h08 3', mRNA sequence.
ACCESSION BJI135240.1 GI:18295397
VERSION EST.
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 588)
AUTHORS Kohara.Y., Shin-I.T., Thierry-Mieg.J., Thierry-Mieg.D., Suzuki.Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
    source
    1..588
    /organism="Caenorhabditis elegans"
    /strain="N2"
    /db_xref="taxon:6239"
    /clone="yk1097h08"
    /clone_lib="unpublished oligo-capped cDNA library, C.
    elegans L1 stage"
    /sex="hermaphrodite"
    /tissue_type="whole animal"
    /dev_stage="L1"
    /note="The AD-wimcDNA library was generated with poly(A)+
    RNA isolated from both hermaphrodite and male N2 worms of
    all larval stages, embryos, adults and dauers and the
    subsequent generation of cDNAs by poly(A) priming. The
    cDNAs were cloned into pPC86"
BASE COUNT 166 a 112 c 131 g 179 t
ORIGIN

Query Match 4.28; Score 45.6; DB 13; Length 588;
Best Local Similarity 59.18; Pred. No. 0.56;
Matches 78; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 191 TCTAAGGAGAACACGACGTATATCAATGAAAGCGGTGTCCTGAGGAGAACTG 250
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 TCAAAATCTACAACTCTCCGGCTTATCTACTCAACGCATATGTAAGCTCTCCAAATCA 320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 251 ATAAGACGGCTCCATATGCTTTCTTCTTCATGTTTTCATTAAGTCGCGCTTCAACA 310
    ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 AAAAGTCGCGCTCCGTAAGCTTTTCTCCCGCATCTTGTGACAGATTCCAAATTCGATT 380
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 311 GTAACACAAAAG 322
    ||| ||| ||| |||
Db 381 GTCAACCAAGTAG 392

```

```

RESULT 13
LOCUS CNS0100X 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098379.1 GI:5609990
VERSION GSS.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Eukaryota, Metazoa, Arthropoda, Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phytophagidae; Drosophilidae; Drosophila.
JOURNAL Genoscope.
COMMENT Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
    Location/Qualifiers
    1..1101
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="BACN03G04"
    /clone_lib="DrosBAC"
    /plasmid="pBelobAC11"
    /note="end : SP6"
BASE COUNT 195 a 108 c 131 g 161 t 506 others
ORIGIN

Query Match 4.28; Score 45.2; DB 17; Length 1101;
Best Local Similarity 17.68; Pred. No. 0.74;
Matches 94; Conservative 212; Mismatches 226; Indels 1; Gaps 1;

Qy 357 TTCTATGCGCTGTTTCTTACTAGGAAGAGCTTGTACTTTTCTATGACTTTAGTGAAGA 416
    : : | | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 WTWTTTTGKKWDTKTKWKDDGGTTWTDTWKDKWTKDKDTKTWDMADNAGWMTOKD 522
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 417 GTCTTCCCATGTTTATGAAAATTCAGAAATGAGGATGTAGAACCAAGCGAGTGCC 476
    : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 523 KTRAKKKRTKTDTKTKRTGTGKKTKDGMKRTDTWGWGDMRADRTDGMADRAANAKDK 582
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 477 CTAAAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATATCTTTATCGAGAGTTCCGA 536
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 583 DADGKADTDATKTDGNGYTTTADTTDDADWDKMDKMDKADTKAGRRGNGWKKGG 642
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 537 TTACAGAGCAATGGGAAAGTAACGGTCTGCAACAGACAGACAAGTATTCATGGGAGCTA 596
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 643 WKKTITKTKKKYDKTKTCTDTKWKMGDKWTKDKKRDGDKRGDGTGGDGSRAKGD 702
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 597 CAATCCTGAAACGGGATAATAGGAAAAATGTCTTCTTAATTCGAAGAAATTAATGACTG 656
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 703 AAWRAKATRAAADAATAATAAATAKATAWATKTTTKTTTTTTTTTTTTTTTTTKTTT 762
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 657 CTGTAGTGGTCTAAGAAATCAGAAAGGAGACCGCAAGAGTCTTAAGTAAATCAAGAAAAAT 716
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 763 ARKADRDDDDKAKRDGDKGDWDADKDWKAGRDRODDTRKW-GWKDRAGKKKKAKKDH 821
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 717 TTGACAAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAGAGCTCGATACCAATAATTC 776
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 KTAWWNADDTGTAWATWAKTATAKTGAKATKDBKRGRRGGGGRGDRGRRKRARGKR 881
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 777 GATTAATAAAGGAGGTGCTTACACAGCCATATCAGTGGAGCAACCTTCCCTTGTGATTCT 836

```


RESULT 18	
BJ402020	
LOCUS	547 bp mRNA linear EST 10-MAR-2002
DEFINITION	Dictyostelium discoideum cDNA library, SF Dictyostelium
ACCESSION	dictyostelium cDNA clone dds20n04 3', mRNA sequence.
VERSION	BJ402020
KEYWORDS	BJ402020.1 GI:19314937
SOURCE	EST.
ORGANISM	Dictyostelium discoideum.
REFERENCE	Dictyostelium discoideum.
AUTHORS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
TITLE	1 (bases 1 to 547)
JOURNAL	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
COMMENT	Full length cDNA of Dictyostelium discoideum at the slug stage
	Unpublished (2002)
	Contact: Tadasu Shin-i

```

FEATURES      source
Email: tshnigenes.nlg.ac.jp.
Location/Qualifiers
1. 362
  organism="Dictyostelium discoideum"
  /bitfield="AY4"
  /clone="taxon:44689"
  /clone="ddc30q10"
  /clone.lib="dictyostelium discoideum"
  /sex="mat A"
  /dev_stage="Culmination stage"
191 a_84 c_93 q_192 t_2 0
BASE COUNT

```

	Query Match	4.1%	Score 45;	DB 13;	Length 562;
	Best Local Similarity	61.5%;	Pred. NO. 0.79;		
	Matches 72;	Conservative	0;	Mismatches 45;	Indels 0;
	Gaps	0;			
QY	208	TACGTTTATCAATGAAGCGTGTCTCAAGTTCTCTGAGGAACTGTATAGAACGGCTCCATA	267		
DB	207	TGCTTTTATCAGTTTAAAGATGTTTCATTTCCACCTGTTTGATGATATAAATACCTGCACCATA	266		


```

source
1. .598
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds17c11"
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="slug stage"
203 a 88 c 207 t 1 others
BASE COUNT 203 a 88 c 207 t 1 others
ORIGIN

Query Match 4.1%; Score 45; DB 13; Length 598;
Best Local Similarity 61.5%; Pred. No. 0.8;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACGTTATCAATGAAGCGGTCTCCAACTTCCTCAGGAGAACTGTAAGAAGCGCTCCATA 267
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 237 TGCCTTTATCAGTTAAAAGAGTGTCCATTTCCACCTGTTGATGATAAAATACCTGCACCAT 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 268 TCGTTTCTTCCTTCATGTTTTCATTAAGTCGCGTTTCCACAGTAAACCAAGA 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 297 TGCCTCAATTGTTATCACCTCTTTACATAATCCAAATTCACAGTCTAAACCAATAACA 353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 25
C93147/c
LOCUS
DEFINITION
C93147 Dictyostelium discoideum ss (H.Urushihara) Dictyostelium
discoideum cDNA clone SSM681, mRNA sequence.
C93147
C93147.1 GI:3075023
EST.
Dictyostelium discoideum.
Dictyostelium discoideum
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
1 (bases 1 to 653)
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka.Y.
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development
DNA Res. 5 (6), 335-340 (1998)
99156227
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan'
POLYA-NO. Location/Qualifiers
1. .653
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSM681"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"
231 a 107 c 217 t
BASE COUNT 231 a 107 c 217 t
ORIGIN

Query Match 4.1%; Score 45; DB 14; Length 653;
Best Local Similarity 61.5%; Pred. No. 0.8;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACGTTATCAATGAAGCGGTCTCCAACTTCCTCAGGAGAACTGTAAGAAGCGCTCCATA 267
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 370 TGCCTTTATCAGTTAAAAGAGTGTCCATTTCCACCTGTTGATGATAAAATACCTGCACCAT 311
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 268 TCGTTTCTTCCTTCATGTTTTCATTAAGTCGCGTTTCCACAGTAAACCAAGA 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Fri Jan 10 12:01:29 2003

```

LOCUS       BJ399159             688 bp    mRNA          linear      EST 10-MAR-2002
DEFINITION  BJ399159 Dictyostellium discoideum cDNA library, SF Dictyostellium
            discoideum cDNA clone dds4p10 3', mRNA sequence.
ACCESSION   BJ399159
VERSION     BJ399159.1  GI:19312076
KEYWORDS    Dictyostellium discoideum.
SOURCE      Dictyostellium discoideum.
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 688)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source           1..688
                     /organism="Dictyostellium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone_lib="dds4p10"
                     /sex="mat A"
     BASE COUNT      231 a 102 c 114 g 240 t 1 others

Query Match      4.1%; Score 45; DB 13; Length 688;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 237 TGCCTTATCAGTAAAGTGTCCATTTCACCTGTGATGATAAATACCTGCACCATTA 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 268 TGCCTTTCTCTTCATGTTTCAATAGTCGCTTTCACAGTAAACCAAAAGCA 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 297 TGCCTAAATGTACCTCTTTTACATATCCAAATTCACCTGTAAACCAATAACA 353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 33
LOCUS       BJ341290             691 bp    mRNA          linear      EST 07-MAR-2002
DEFINITION  BJ341290 Dictyostellium discoideum cDNA library, AF Dictyostellium
            discoideum cDNA clone dds6d03 3', mRNA sequence.
ACCESSION   BJ341290
VERSION     BJ341290.1  GI:19249652
KEYWORDS    Dictyostellium discoideum.
SOURCE      Dictyostellium discoideum.
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 691)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the aggregation
            stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source           1..691
                     /organism="Dictyostellium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"

```

```

/clone="dds6d03"
/clone_lib="Dictyostellium discoideum cDNA library, AF"
/sex="mat A"
/dev_stage="Aggregation stage"
BASE COUNT      229 a 101 c 121 g 238 t 2 others

Query Match      4.1%; Score 45; DB 13; Length 691;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 185 TGCCTTATCAGTAAAGTGTCCATTTCACCTGTGATGATAAATACCTGCACCATTA 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 268 TGCCTTTCTCTTCATGTTTCAATAGTCGCTTTCACAGTAAACCAAAAGCA 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 245 TGCCTAAATGTATCACCCTCTTTTACATATCCAAATTCACCTGTAAACCAATAACA 301
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 34
LOCUS       BJ401883             698 bp    mRNA          linear      EST 10-MAR-2002
DEFINITION  BJ401883 Dictyostellium discoideum cDNA library, SF Dictyostellium
            discoideum cDNA clone dds19p14 3', mRNA sequence.
ACCESSION   BJ401883
VERSION     BJ401883.1  GI:19314800
KEYWORDS    Dictyostellium discoideum.
SOURCE      Dictyostellium discoideum.
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 698)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source           1..698
                     /organism="Dictyostellium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="dds19p14"
                     /clone_lib="Dictyostellium discoideum cDNA library, SF"
                     /sex="mat A"
     BASE COUNT      231 a 103 c 119 g 244 t 1 others

Query Match      4.1%; Score 45; DB 13; Length 698;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 269 TGCCTTATCAGTAAAGTGTCCATTTCACCTGTGATGATAAATACCTGCACCATTA 328
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 268 TGCCTTTCTCTTCATGTTTCAATAGTCGCTTTCACAGTAAACCAAAAGCA 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 329 TGCCTAAATGTATCACCCTCTTTTACATATCCAAATTCACCTGTAAACCAATAACA 385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 35
LOCUS       BJ400818             718 bp    mRNA          linear      EST 10-MAR-2002
DEFINITION  BJ400818 Dictyostellium discoideum cDNA library, SF Dictyostellium
            discoideum cDNA clone dds14p23 3', mRNA sequence.
ACCESSION   BJ400818

```

```

VERSION      BJ400818.1  GI:19313735
KEYWORDS     EST.
SOURCE       Dictyostellium discoideum.
ORGANISM     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE    1 (bases 1 to 718)
AUTHORS      Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE        Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL      Unpublished (2002)
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.

FEATURES     Source
             1..718
             /organism="Dictyostellium discoideum"
             /strain="AX4"
             /db_xref="taxon:44689"
             /clone_lib="ddas2208"
             /clone_lib="Dictyostellium discoideum cDNA library, SF"
             /sex="mat A"
             /dev_stage="Slug stage"
BASE COUNT   241 a 104 c 122 g 248 t 3 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 718;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCTCGAGGAGAACTGATAAGACGGCTCCATA 267
DB 198 TGTCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAATACCTGCACCATTA 257
QY 268 TGCCTTTCTTCCTTCATGTTTTCATTAAGTCGCGTTTCAACAGTAAACCAAGCA 324
DB 258 TGCTCTAATTGTATCACCTTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 314

RESULT 36
LOCUS       BJ401195
DEFINITION Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION   BJ401195
VERSION     BJ401195.1  GI:19314112
KEYWORDS    EST.
SOURCE      Dictyostellium discoideum.
ORGANISM    Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 722)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
              Location/Qualifiers
                1..722
                /organism="Dictyostellium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="ddas2208"
                /clone_lib="Dictyostellium discoideum cDNA library, AF"
                /sex="mat A"
                /dev_stage="Aggregation stage"
BASE COUNT  243 a 105 c 122 g 249 t 3 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 724;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCTCGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGCCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAATACCTGCACCATTA 296
QY 268 TGCCTTTCTTCCTTCATGTTTTCATTAAGTCGCGTTTCAACAGTAAACCAAGCA 324
DB 297 TGCTCTAATTGTATCACCTTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 353

RESULT 38
LOCUS       BJ343890
DEFINITION Dictyostellium discoideum cDNA library, AF Dictyostellium
ACCESSION   BJ343890
VERSION     BJ343890.1  GI:19214397
KEYWORDS    EST.
SOURCE      Dictyostellium discoideum.
ORGANISM    Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.
REFERENCE   1 (bases 1 to 734)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE       Full length cDNA of Dictyostellium discoideum at the aggregation
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshini@genes.nig.ac.jp.
              Location/Qualifiers
                1..734
                /organism="Dictyostellium discoideum"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone_lib="ddas120"
                /clone_lib="Dictyostellium discoideum cDNA library, AF"
                /sex="mat A"
                /dev_stage="Aggregation stage"
BASE COUNT  243 a 106 c 124 g 251 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 724;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCTCGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGCCTTATCAAGTAAAGAGTGTCCATTTACCTGTGATGATAAATACCTGCACCATTA 296
QY 268 TGCCTTTCTTCCTTCATGTTTTCATTAAGTCGCGTTTCAACAGTAAACCAAGCA 324
DB 297 TGCTCTAATTGTATCACCTTCTTTACATAATCCAAATTCAACTGTAACCAATAACA 353

```



```

Eukaryota; Mycetozoa; Dictyostelidia; Dictyostellium.
1 (bases 1 to 734)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostellium discoideum at the aggregation
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
    source
    1..734
    /organism="Dictyostellium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone_lib="ddal7d13"
    /clone_lib="Dictyostellium discoideum cDNA library, AF"
    /sex="mat A"
    /dev_stage="Aggregation stage"
    /dev_stage="Aggregation stage"

BASE COUNT 247 a 105 c 126 g 256 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 734;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TCGTTTATCAGTTAAAAGTGTTCATTTTCACTGTGTGATGATAAAATACCTGCACCATA 295
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
QY 268 TCGTTTCTTCCTTCATGCTGTTTCAATAAGTTCGCTTTCAACAGTAAACCAAAAGCA 324
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCCTAATTTGATCACCTTCTTTACATAATCCAATTCACCTGTAAACCAATAACA 352
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

FEATURES
    source
    1..734
    /organism="Dictyostellium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone_lib="ddal7d13"
    /clone_lib="Dictyostellium discoideum cDNA library, AF"
    /sex="mat A"
    /dev_stage="Aggregation stage"
    /dev_stage="Aggregation stage"

BASE COUNT 247 a 105 c 126 g 256 t
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 734;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TCGTTTATCAGTTAAAAGTGTTCATTTTCACTGTGTGATGATAAAATACCTGCACCATA 295
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
QY 268 TCGTTTCTTCCTTCATGCTGTTTCAATAAGTTCGCTTTCAACAGTAAACCAAAAGCA 324
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCCTAATTTGATCACCTTCTTTACATAATCCAATTCACCTGTAAACCAATAACA 352
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

FEATURES
    source
    1..742
    /organism="Dictyostellium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone_lib="dd5c06"
    /clone_lib="Dictyostellium discoideum cDNA library, SF"
    /sex="mat A"
    /dev_stage="Slug stage"
    /dev_stage="Slug stage"

BASE COUNT 249 a 105 c 126 g 261 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 742;

```

```

Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TCGTTTATCAGTTAAAAGTGTTCATTTTCACTGTGTGATGATAAAATACCTGCACCATA 295
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
QY 268 TCGTTTCTTCCTTCATGCTGTTTCAATAAGTTCGCTTTCAACAGTAAACCAAAAGCA 324
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCCTAATTTGATCACCTTCTTTACATAATCCAATTCACCTGTAAACCAATAACA 352
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

RESULT 40
BJ347280 744 bp mRNA linear EST 06-MAR-2002
LOCUS BJ347280 Dictyostellium discoideum cDNA library, AF Dictyostellium
DEFINITION BJ347280 Dictyostellium discoideum cDNA clone dda26k16 3', mRNA sequence.
ACCESSION BJ347280
VERSION BJ347280.1 GI:19217787
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum
REFERENCE 1 (bases 1 to 744)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostellium discoideum at the aggregation
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..744
/organism="Dictyostellium discoideum"
/db_xref="taxon:44689"
/clone_lib="dd26k16"
/clone_lib="Dictyostellium discoideum cDNA library, AF"
/sex="mat A"
/dev_stage="Aggregation stage"
/dev_stage="Aggregation stage"

BASE COUNT 249 a 105 c 127 g 262 t 1 others
ORIGIN
Query Match 4.1%; Score 45; DB 13; Length 744;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TAGCTTATCAATGAAGCGTGTCCAAAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 236 TCGTTTATCAGTTAAAAGTGTTCATTTTCACTGTGTGATGATAAAATACCTGCACCATA 295
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
QY 268 TCGTTTCTTCCTTCATGCTGTTTCAATAAGTTCGCTTTCAACAGTAAACCAAAAGCA 324
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||
Db 296 TGCCTAATTTGATCACCTTCTTTACATAATCCAATTCACCTGTAAACCAATAACA 352
    1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 |||||

RESULT 41
BJ401342 760 bp mRNA linear EST 10-MAR-2002
LOCUS BJ401342 Dictyostellium discoideum cDNA library, SF Dictyostellium
DEFINITION BJ401342 Dictyostellium discoideum cDNA clone dds22019 3', mRNA sequence.
ACCESSION BJ401342
VERSION BJ401342.1 GI:19314259
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum
REFERENCE 1 (bases 1 to 760)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.

```

TITLE Full length cDNA of Dictyostellium discoideum at the slug stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..760
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds22019"
/clone_lib="Dictyostellium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"
257 a 109 c 127 g 266 t 1 others
BASE COUNT
ORIGIN

Query Match 4.1%; Score 45; DB 13; Length 760;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACCTTATCAATGAAGCGTGTCCAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGCCTTATCAAGTAAAAAGTGTCCATTCACCTGTGTGATGATAAAATACCTGCACCATTA 296
QY 268 TGCCTTCTCTCCCTTCATGTTTCAATAGTCGCTTTCACAGTAAACCAACCAAGCA 324
DB 297 TGCCTTAATGTATCACCTCTTACATAATCCAAATTCACCTGTAACCAATAACA 353

RESULT 42
B371948
LOCUS B371948 Dictyostellium discoideum cDNA library, CF Dictyostellium
DEFINITION dictyostellium cDNA clone ddc10p12 3', mRNA sequence.
ACCESSION B371948
VERSION B371948.1 GI:19281331
KEYWORDS EST.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 765)
AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
TITLE Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the culmination stage

JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES
source
1..765
/organism="Dictyostellium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc10p12"
/clone_lib="Dictyostellium discoideum cDNA library, CF"
/sex="mat A"
/dev_stage="Culmination stage"
258 a 113 c 126 g 268 t
BASE COUNT
ORIGIN

Query Match 4.1%; Score 45; DB 13; Length 765;
Best Local Similarity 61.5%; Pred. No. 0.81;
Matches 72; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 208 TACCTTATCAATGAAGCGTGTCCAGTTCCTGAGGAGAACTGATAAGACGGCTCCATA 267
DB 237 TGCCTTATCAAGTAAAAAGTGTCCATTCACCTGTGTGATGATAAAATACCTGCACCATTA 296
QY 268 TGCCTTCTCTCCCTTCATGTTTCAATAGTCGCTTTCACAGTAAACCAACCAAGCA 324
DB 297 TGCCTTAATGTATCACCTCTTACATAATCCAAATTCACCTGTAACCAATAACA 353

RESULT 43
CNS0072R
LOCUS CNS0072R Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR14E13 of RPCI-98 library from Drosophila melanogaster (fruit fly) genomic survey sequence.
ACCESSION AL066743
VERSION AL066743.1 GI:4945206
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 988)
AUTHORS Direct Submission
TITLE Genoscope.
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> the BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's Pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..988
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR14E13"
/clone_lib="RPCI-98"
/note="end : TET3"
BASE COUNT 462 a 175 c 86 g 159 t 106 others
ORIGIN

Query Match 4.1%; Score 44.6; DB 17; Length 988;
Best Local Similarity 39.3%; Pred. No. 1;
Matches 107; Conservative 31; Mismatches 134; Indels 0; Gaps 0;

QY 623 AAATGTCTTTCTAATTGCGAAGATTTAATGACTGCTGTGATGCTCTAAGAAATCAGAAAGG 682
DB 540 AACTYATTCACCAAAAWWWAAAAAAMWWTAAAAAATAAACCTTAAAAAATAAAAAA 599
QY 683 AGACCGAAGCTTCTAAGTAATCAAGAAAAAATCTTGACAAATAGCTTTTCCATAGAGAA 742
DB 600 AACCTAAAHTCTCAAAAAAATAAAAAAATCTCTYAAAAAAMCTTTCMACAYACAAM 659
QY 743 AACCGTGAGGAAGGAGTCGATACCAATAAATCCGATTAATAAAGGAGCTTACCAACA 802
DB 660 WTMAAAWHCCAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA 719
QY 803 GCCATATCAGTGGGACCACTCCCTTTGATCTTTGATATGACAGCAAGATCAACAAG 862
DB 720 YTTATWTCYTAATYTWATATCAAAATWTHNTCAHAATAAATAAAAAAATAAAAAAATA 779

RESULT 46

AZ550349
LOCUS
DEFINITION
ENTEM79TF Entamoeba histolytica sheared DNA linear GSS 14-NOV-2000
genomic, DNA sequence.
ACCESSION
AZ550349
VERSION
AZ550349.1 GI:11175650
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica.
REFERENCE
1 (bases 1 to 896)
AUTHORS
Loftus B., Van Aken S. and Fraser C.
TITLE
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL
Unpublished (2000)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 36
High quality sequence stop: 844.
FEATURES
source
1..896
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 464 a 66 c 161 g 205 t
ORIGIN
Query Match 4.0%; Score 43.8; DB 17; Length 896;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 532 TCGCATTTACAGAGCAATGGGAAAGTACGGTCTCGAACAGACAGCAAGTATTGATGGG 591
Db 430 TCAAAATACAAAAGAAATTTATTTCAATTTATTTGTTGAAAAACAAAAGAAATTTAAAGGAA 489
Qy 592 AGCTACAAATCTCGAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAAT 651
Db 490 TTTTAAAGAAGATGGCAACAAGAAAATAGAAACATGAATTAATTAATTAAGAAAGGAA 549
Qy 652 GACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCGCAAGCTTCTTAAGTAATCAAGAAA 711
Db 550 GAAAAATGGATAGATTAAAAAACAAGAAACAATGGAAACACAAAAGATGGAGAA 609
Qy 712 AAA 714
Db 610 AAA 612

RESULT 47
AZ527921/c
LOCUS
DEFINITION
ENTCG177R Entamoeba histolytica sheared DNA linear GSS 03-NOV-2000
genomic, DNA sequence.
ACCESSION
AZ527921
VERSION
AZ527921.1 GI:11080094
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica.
REFERENCE
1 (bases 1 to 899)
AUTHORS
Loftus B., Van Aken S. and Fraser C.
TITLE
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL
Unpublished (2000)
COMMENT
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 808.
FEATURES
source
1..899
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 208 a 151 c 64 g 476 t
ORIGIN
Query Match 4.0%; Score 43.8; DB 17; Length 899;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 532 TCGCATTTACAGAGCAATGGGAAAGTACGGTCTCGAACAGACAGCAAGTATTGATGGG 591
Db 554 TCAAAATACAAAAGAAATTTATTTCAATTTATTTGTTGAAAAACAAAAGAAATTTAAAGGAA 495
Qy 592 AGCTACAAATCTCGAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAATTTAAT 651
Db 494 TTTTAAAGAAGATGGCAACAAGAAAATAGAAACATGAATTAATTAATTAAGAAAGGAA 435
Qy 652 GACTGCTTGATGCTCTAAGAAATCAGAAAGGAGACCGCAAGCTTCTTAAGTAATCAAGAAA 711
Db 434 GAAAAATGGATAGATTAAAAAACAAGAAACAATGGAAACACAAAAGATGGAGAA 375
Qy 712 AAA 714
Db 374 AAA 372

RESULT 48
 A2687856/c
 LOCUS ENTL72TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, DNA sequence.
 ACCESSION A2687856
 VERSION A2687856.1 GI:11825002
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 914)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library
 COMMENT Unpublished (2000)
 CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 7
 High quality sequence stop: 869.
 Location/Qualifiers
 1. 914
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaubin and B.
 Barrel, Oxford University Press, 1999)."

FEATURES

Source
 1. 914

BASE COUNT 209 a 155 c 85 g 465 t
 ORIGIN
 Query Match 4.0%; Score 43.8; DB 17; Length 914;
 Best Local Similarity 52.5%; Pred. No. 1.7;
 Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 532 TCGCATTACAGAGGCAATGGAAAGTATTCATTAATTCGAAAGAAATTTAAT 651
 Db 260 TCAAAATACAAAAGAAATTTATTTCAATTAATTCGAAAGAAAGAAATTTAAGGAA 201
 QY 592 AGCTACAATCCTCAAGGGAATATAGGAAATGTTTCTTAATTCGAAAGAAATTTAAT 651
 Db 200 TTTTAAAGAAGATGGAACAAAGAAATAGAAACATGAATTAATTAATTAAGAAAGGAA 141
 QY 652 GACTGCTTGTGCTTAAGAATCAGAAAGGAGACCGGAAGCTTCTAAGTAATCAAGAAA 711
 Db 140 GAAAAATGGATAGATTAAAAAACAAGAAACAAATGGAACACACAAAAGATGGAGAA 81
 QY 712 AAA 714
 Db 80 AAA 78

RESULT 49

A2690917
 LOCUS ENTHV65TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, DNA sequence.
 ACCESSION A2690917 GI:11828063
 VERSION A2690917.1
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.
 ORGANISM Entamoeba histolytica.
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 924)
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 JOURNAL HMI:IMSS sheared DNA library
 COMMENT Unpublished (2000)
 CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjlloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 85
 High quality sequence stop: 895.
 Location/Qualifiers
 1. 924
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaubin and B.
 Barrel, Oxford University Press, 1999)."

BASE COUNT

482 a 63 c 169 g 210 t

ORIGIN

Query Match 4.0%; Score 43.8; DB 17; Length 924;
 Best Local Similarity 52.5%; Pred. No. 1.7;
 Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 532 TCGCATTACAGAGGCAATGGAAAGTATTCATTAATTCGAAAGAAATTTAAT 651
 Db 116 TCAAAATACAAAAGAAATTTATTTCAATTAATTCGAAAGAAAGAAATTTAAGGAA 175
 QY 592 AGCTACAATCCTCAAGGGAATATAGGAAATGTTTCTTAATTCGAAAGAAATTTAAT 651
 Db 176 TTTTAAAGAAGATGGAACAAAGAAATAGAAACATGAATTAATTAATTAAGAAAGGAA 235
 QY 652 GACTGCTTGTGCTTAAGAATCAGAAAGGAGACCGGAAGCTTCTAAGTAATCAAGAAA 711
 Db 236 GAAAAATGGATAGATTAAAAAACAAGAAACAAATGGAACACACAAAAGATGGAGAA 295
 QY 712 AAA 714
 Db 296 AAA 298

RESULT 50
 AZ549336

LOCUS AZ549336 944 bp DNA linear GSS 14-NOV-2000
DEFINITION ENTED22TRB Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION AZ549336
VERSION AZ549336.1 GI:11173846
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 944)
Lofthus, S., Van Aken, S. and Fraser, C. C. Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library
Unpublished (2000)
JOURNAL
COMMENT Contact: Brendan J Lofthus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library

Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 907.
FEATURES
source
1..944
Location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="vector: pHOsl; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 500 a 60 c 162 g 222 t
ORIGIN
Query Match 4.08; Score 43.8; DB 17; Length 944;
Best Local Similarity 52.5%; Pred. NO. 1.7;
Matches 96; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
Qy 532 TCGCATACAGAGCGCAATGGGAAGTAAACGGCTCTGCAACAGACAGACAAGTATTGATGGG 591
Dy 376 TCAATTTACAAAAGAAATTTTTCATTTTGTGAAAACAAAAGAAATTTAAAGGAA 435
Qy 592 AGCTACAAATCTGAACGGGATATAGGAAAATGTGTTCTTAATTCGAAGAATTTAAT 651
Dy 436 TTTTAAAGAAGATGGAACAAGAAATAGAAACATGAATTAATTAATAAAGAAAGGAA 495
Qy 652 GACTGCTTCTTAAGAAATCAGAAAGAGACCGGAAAGCTTCTAAGTAATCAAGAAA 711
Dy 496 GAAATGATAGATTAAAAAACAAAAGAAACAAATGGAACACAAAGATGGAGAA 555
Qy 712 AAA 714
Dy 556 AAA 558

Search completed: January 9, 2003, 11:40:55
Job time: 1961 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:32:32 ; Search time 251 Seconds

(without alignments)

9743.706 Million cell updates/sec

Title: us-09-438-185a-1_COPY_1199590_1200675

Perfect score: 1086

Sequence: 1 ttggcaagtagtacctcaaac.....gtctctgcgtagtgac 1086

Scoring table: IDENTITY NUC
Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

N_Geneseq_101002:*

- 1: /SID22/gcgdata/geneq/geneq-emb1/NA1980.DAT:*
- 2: /SID22/gcgdata/geneq/geneq-emb1/NA1981.DAT:*
- 3: /SID22/gcgdata/geneq/geneq-emb1/NA1982.DAT:*
- 4: /SID22/gcgdata/geneq/geneq-emb1/NA1983.DAT:*
- 5: /SID22/gcgdata/geneq/geneq-emb1/NA1984.DAT:*
- 6: /SID22/gcgdata/geneq/geneq-emb1/NA1985.DAT:*
- 7: /SID22/gcgdata/geneq/geneq-emb1/NA1986.DAT:*
- 8: /SID22/gcgdata/geneq/geneq-emb1/NA1987.DAT:*
- 9: /SID22/gcgdata/geneq/geneq-emb1/NA1988.DAT:*
- 10: /SID22/gcgdata/geneq/geneq-emb1/NA1989.DAT:*
- 11: /SID22/gcgdata/geneq/geneq-emb1/NA1990.DAT:*
- 12: /SID22/gcgdata/geneq/geneq-emb1/NA1991.DAT:*
- 13: /SID22/gcgdata/geneq/geneq-emb1/NA1992.DAT:*
- 14: /SID22/gcgdata/geneq/geneq-emb1/NA1993.DAT:*
- 15: /SID22/gcgdata/geneq/geneq-emb1/NA1994.DAT:*
- 16: /SID22/gcgdata/geneq/geneq-emb1/NA1995.DAT:*
- 17: /SID22/gcgdata/geneq/geneq-emb1/NA1996.DAT:*
- 18: /SID22/gcgdata/geneq/geneq-emb1/NA1997.DAT:*
- 19: /SID22/gcgdata/geneq/geneq-emb1/NA1998.DAT:*
- 20: /SID22/gcgdata/geneq/geneq-emb1/NA1999.DAT:*
- 21: /SID22/gcgdata/geneq/geneq-emb1/NA2000.DAT:*
- 22: /SID22/gcgdata/geneq/geneq-emb1/NA2001A.DAT:*
- 23: /SID22/gcgdata/geneq/geneq-emb1/NA2001B.DAT:*
- 24: /SID22/gcgdata/geneq/geneq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1086	100.0	1089	ABL91284	Chlamydia pneumoniae
C 2	46.6	4.3	1642	AAA91624	Caenorhabditis ele
C 3	43.4	4.0	6089	ABN80254	Human chemically m
C 4	41.4	3.8	817	AAA91625	Caenorhabditis ele
C 5	41	3.8	868	AA217250	Human gene express
6	41	3.8	9504	ABK28408	DNA transcription
7	40.6	3.7	5152	ABL92307	Chemically treated
8	40.6	3.7	5152	ABL49374	Human polynucleoti
9	40.6	3.7	37184	ABQ67077	Human angiogenesis

10	40.6	3.7	640681	24	ABA92787	Buchnera sp. genom
11	40	3.7	6740	24	ABL33147	Human immune syste
12	39.6	3.6	17419	22	AAS45393	Chemically pretrea
13	39.6	3.6	17419	24	ABL33295	Human immune syste
14	39.6	3.6	17419	24	ABK28238	DNA transcription
C 15	39.6	3.6	99629	22	ABK28550	Genomic fragment #
C 16	39.6	3.6	8895	24	ABL32333	Human immune syste
C 17	38.6	3.6	326	23	ABV61126	Human prostate exp
18	38.6	3.5	18997	24	ABK33948	Human DNA for stag
19	38.4	3.5	18997	24	ABL32570	Human immune syste
C 20	38.4	3.5	50000	24	ABK26437	Human glutamate re
C 21	38.4	3.5	83391	24	ABQ67094	Human GRM3 gene fr
C 22	38.4	3.5	7657	22	AAS46333	Human angiogenesis
23	38.2	3.5	9415	24	ABL32294	Tumour suppressor
C 24	38.2	3.5	15373	24	ABL32467	Human immune syste
C 25	38.2	3.5	47108	24	ABK31510	Signal transductio
C 26	38.2	3.5	17213	24	ABL33483	Human immune syste
C 27	38	3.5	113515	24	ABL34174	Human immune syste
C 28	38	3.5	7690	24	ABL33122	Arabidopsis thalia
C 29	37.8	3.5	870	21	AAC40730	Human immune syste
C 30	37.4	3.4	6237	24	ABL32359	Human immune syste
31	37.4	3.4	7644	24	ABL32530	Human immune syste
C 32	37.2	3.4	522	21	AAC36518	Arabidopsis thalia
C 33	37.2	3.4	869	21	AAC51121	Arabidopsis thalia
C 34	37.2	3.4	3152	22	AAH54692	S. epidermidis gen
35	37.2	3.4	3345	22	AAH54004	S. epidermidis gen
36	37.2	3.4	6113	24	ABL32802	Human immune syste
37	37.2	3.4	7190	24	ABL33720	Human immune syste
C 38	37	3.4	619	23	ABV52129	Human prostate exp
C 39	37	3.4	1134	20	AAZ31005	Partial dnan gene.
40	37	3.4	1075	22	AAH34735	Nucleotide sequenc
41	37	3.4	1075	22	AAH32611	Human secreted pro
C 42	36.8	3.4	1724	22	AAK32609	Human immune/hema
C 43	36.8	3.4	1761	22	AAK68427	Human immune/hema
44	36.8	3.4	1761	22	AAK68428	Human angiogenesis
45	36.8	3.4	7857	24	ABQ67075	Listeria monocytog
C 46	36.8	3.4	294528	24	ABX03041	Human cDNA clone (
C 47	36.8	3.4	771	12	AAH07351	Homo sapiens solub
C 48	36.6	3.4	2098	19	AAV26649	Human cDNA sequenc
C 49	36.6	3.4	2333	22	AAH17997	Human immune syste
C 50	36.6	3.4	5276	24	ABL32151	Tumour suppressor
51	36.6	3.4	7348	22	AAS46335	Arabidopsis thalia
C 52	36.6	3.4	1241	22	ABD09807	Chemically treated
53	36.4	3.4	6239	24	ABL70584	Human gene regulat
C 54	36.4	3.4	6239	24	AAS61287	Signal transductio
C 55	36.4	3.4	8962	24	ABL31387	Human immune syste
C 56	36.4	3.4	15282	24	ABL32687	Chemically treated
57	36.4	3.4	15282	24	ABL70189	Human gene regulat
58	36.4	3.4	21354	22	AAS61139	Human gene regulat
59	36.4	3.4	21354	22	AAS46815	Tumour suppressor
60	36.4	3.4	319630	24	ABQ67194	Listeria innocua c
C 61	36.4	3.4	373	23	ABV60277	Human prostate exp
C 62	36.2	3.3	444	22	AAH69056	Human cervical can
63	36.2	3.3	470	22	AAH71919	Human cervical can
C 64	36.2	3.3	511	22	AAH70719	Human cervical can
65	36.2	3.3	580	24	ABK44560	cDNA encoding colo
C 66	36.2	3.3	585	24	ABN66198	Streptococcus poly
C 67	36.2	3.3	1985	20	AAZ32066	Human METH2 relate
C 68	36.2	3.3	1985	22	AAC90323	Human histone deac
C 69	36.2	3.3	1985	22	AAC89555	Human cDNA encodin
C 70	36.2	3.3	1985	22	ABK87716	Human polynucleoti
C 71	36.2	3.3	6020	24	ABL49308	Human immune syste
72	36.2	3.3	6020	24	ABL32199	Human angiogenesis
73	36.2	3.3	6070	24	ABQ67129	Chemically treated
C 74	36.2	3.3	6070	24	ABL70371	Human immune syste
C 75	36.2	3.3	6070	24	ABL33678	Human metastasis a
C 76	36.2	3.3	6070	24	ABL34578	Chemically treated
C 77	36.2	3.3	10988	24	ABL70200	Human gene regulat
78	36.2	3.3	10988	24	AAS61158	Signal transductio
C 79	36.2	3.3	10988	24	ABK31245	Human immune syste
80	36.2	3.3	17137	22	ABL32191	Human angiogenesis
81	36.2	3.3	18998	22	AAS36452	Human cardiavascu
82	36.2	3.3	18998	22	AAS36452	Human cardiavascu

	Matches	1086:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	1	TTGCAAGTACCTCAAAACCAAGAAATATTTCTCTGATTGTAAAGGGAATGTATTC	60							
Db	1086	TTGCAAGTACCTCAAAACCAAGAAATATTTCTCTGATTGTAAAGGGAATGTATTC	1027							
QY	61	TAAAGACCTTGTGTCAGATCCATCTTAATTTTGAAGTGAGTTCACAGTTCATCAAA	120							
Db	1026	TAAAGACCTTGTGTCAGATCCATCTTAATTTTGAAGTGAGTTCACAGTTCATCAAA	967							
QY	121	ATGTCTTATTGAAAATAAAGTCTCTGTGGAGTTGATGATTGAAGGAAGACCAATAAT	180							
Db	966	ATGTCTTATTGAAAATAAAGTCTCTGTGGAGTTGATGATTGAAGGAAGACCAATAAT	907							
QY	181	CTGATCCAAATCTAAAGGGAGAACACGTACGTTATCAATGAAGCGTGTCCAAAGTCCCTG	240							
Db	906	CTGATCCAAATCTAAAGGGAGAACACGTACGTTATCAATGAAGCGTGTCCAAAGTCCCTG	847							
QY	241	AGGAGAACTGATTAAGAACGGCTCCATATGCTTTTCTTCCTTCATGTTTTCAATAAGTCC	300							
Db	845	AGGAGAACTGATTAAGAACGGCTCCATATGCTTTTCTTCCTTCATGTTTTCAATAAGTCC	787							
QY	301	GCCTTCAACAGTAAACCAAAAGCAGCTACAAATAGCGATCAGATTCGTTGTAGGGTTTG	360							
Db	786	GCCTTCAACAGTAAACCAAAAGCAGCTACAAATAGCGATCAGATTCGTTGTAGGGTTTG	727							
QY	361	TATGGGTGTTTTTACTAGGAAGAGCTTGATCTTTTCTATGACGTTTGTAGAGAGTCT	420							
Db	726	TATGGGTGTTTTTACTAGGAAGAGCTTGATCTTTTCTATGACGTTTGTAGAGAGTCT	667							
QY	421	TCCCATGTTATGAAAAATTCAGAAAAATGAGGATGTAGAAAGCCAAAGCCAGTGCCTAA	480							
Db	666	TCCCATGTTATGAAAAATTCAGAAAAATGAGGATGTAGAAAGCCAAAGCCAGTGCCTAA	607							
QY	481	AAGGTCATGAGTATGAGTACGAGGTTAAGGAGAAATATCTTTATCGAGAGTTCGCATTAC	540							
Db	606	AAGGTCATGAGTATGAGTACGAGGTTAAGGAGAAATATCTTTATCGAGAGTTCGCATTAC	547							
QY	541	AGAGGCAATGGGAAAGTAAACGGCTCTGCAACAGAGACAAGTATTGTAIGGGGAGCTACAAA	600							
Db	546	AGAGGCAATGGGAAAGTAAACGGCTCTGCAACAGAGACAAGTATTGTAIGGGGAGCTACAAA	487							
QY	601	TCTGAAACGGGATTAATGAGAAAAATGTCTTCTAAATCGAAGAAATTTAATGACATGCTTG	660							
Db	486	TCTGAAACGGGATTAATGAGAAAAATGTCTTCTAAATCGAAGAAATTTAATGACATGCTTG	427							
QY	661	ATGGTCTTAAGAAATCAGAAAGGAGACCGCAAGCTTCTAAGTAAATCAAGAAAAAATCTTTGG	720							
Db	426	ATGGTCTTAAGAAATCAGAAAGGAGACCGCAAGCTTCTAAGTAAATCAAGAAAAAATCTTTGG	367							
QY	721	ACAATAGCTTTTCCATAGCAAAACCGTCAGCAAGAGAGTCGATACCATATAATCCGATT	780							
Db	366	ACAATAGCTTTTCCATAGCAAAACCGTCAGCAAGAGAGTCGATACCATATAATCCGATT	307							
QY	781	AAAAAGGAGTGCTTACACACCCATATCAGTGGAGCAACCTTCCTTTGATCTTTTGA	840							
Db	306	AAAAAGGAGTGCTTACACACCCATATCAGTGGAGCAACCTTCCTTTGATCTTTTGA	247							
QY	841	TAATGACAGAGATGAACAAAGTAGTAAGTTTTTGAANAATCTAAATAGAAATGCAATTT	900							
Db	246	TAATGACAGAGATGAACAAAGTAGTAAGTTTTTGAANAATCTAAATAGAAATGCAATTT	187							
QY	901	GTGTCAGACTAAGCTTGCTCTCTTTATTTTTCCTTTTGTGTAAGATGATCGGTAGTAGA	960							
Db	186	GTGTCAGACTAAGCTTGCTCTCTTTATTTTTCCTTTTGTGTAAGATGATCGGTAGTAGA	127							
QY	961	ATATGGGGTCGAGTATGACGTTGGAGTGATTGGCTGTCTTGGGAAGCAACAGGAAGTCA	1020							
Db	126	ATATGGGGTCGAGTATGACGTTGGAGTGATTGGCTGTCTTGGGAAGCAACAGGAAGTCA	67							
QY	1021	TTGTCTCAGCTTTAGAGCAATCTTCAGATATATCTTTGGGTCCAGGGTTCTCTCGCAGTA	1080							
Db	66	TTGTCTCAGCTTTAGAGCAATCTTCAGATATATCTTTGGGTCCAGGGTTCTCTCGCAGTA	7							

Db 1464 TCTTCAAAATTTCTAGTATAGAAATACGCTGACGTGGAAGAGTATGTAATGAGACATCTTTGC 1405
 Qy 173 CGAATATCTGATCCAAATTTCTAAAGGGAGAACAGTACGTTATCAATGAAGCGGTGCCA 232
 Db 1404 TCAACACAGATCCGGATTCAAAGGAATAATGTTGCACCTACCTCAACGCGATGTTGC 1345
 Qy 233 AGTTCTGTGAGGAAGTATGATGAAGACGCTCCATATGCTTTCTCTCTTCATGTTTCA 292
 Db 1344 AACTCGGCAGACTGTCAGAGTCTGCTCGGTATATCTTTAAATCTTTCATGATTTGAT 1285
 Qy 293 ATAAGTCGCTTTCACAGCTAAACAAAGCAGCGGTACATACGATCAGATGCTTTGT 352
 Db 1284 CCAATTTCTTTACTGGAG-RATCGCAGCGTCATCAGACGAGACCAATCAATGGA 1226
 Qy 353 AGGTTTGTATGCTGTTTCTTACTAGGAGAGCTTGTACTTTTCTATGACTTTAGTG 412
 Db 1225 AAGAAGTAGAGTGTGCAAGCTTCTTCAA------TCTTCTCTGATGCTCCAGA 1174
 Qy 413 AAGAGTCTCCCATGTTTATGAATTCAGAAATCAGAAATGAGGATGTAGAACCCAGCAG 472
 Db 1173 GAAGTAACTCAATCTCTTGAGAACTGAGCAAAATCTGATCAGCGATAGACCCATG 1114
 Qy 473 TGCCCTAAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 532
 Db 1113 TGACCATGAGTCTGGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1054
 Qy 533 CGCATTCAGAGGCAATGGGAAAGTAACGCTGCTGCAACAGACAGATGATGATGGGA 592
 Db 1053 CGAACGCTATTGAGTGCAGAGAGACACGATATGCAAGACCTGCCAAGAAATCAGAGCT 994
 Qy 593 GCTCAATCTCTGAAACGGGATATAGGAAATATGTTTCTAATTCGAGAAATTTAATG 652
 Db 993 GATAAGTATCCGCGACTGGCGGACACGGAATCCACTTTTCTTTTCAAAACCTTGCAG 934
 Qy 653 ACTGCTT 659
 Db 933 ATATCTT 927

RESULT 3
 ABN80254/C
 ID ABN80254 standard; DNA; 6089 BP.
 AC ABN80254;
 DT 15-JUL-2002 (first entry)
 XX Human chemically modified disease associated gene SEQ ID NO 271.
 XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KW antidiabetic; cytostatic; anticonvulsant; ds.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200200927-A2.
 XX 03-JAN-2002.
 PD 02-JUL-2001; 2001WO-EP07536.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPTGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K;
 PI WPT; 2002-130908/17.
 XX Novel nucleic acid useful for diagnosis and therapy of diseases

PT associated with development genes such as diabetes, comprises a
 PT sequence of a segment of chemically pretreated DNA of genes associated
 PT with development -
 XX Claim 1: SEQ ID NO 271; 27pp; English.
 XX The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)
 CC of genes associated with development selected from 87 genes listed in
 CC the specification such as ACCPN, ADFN, or AFDI and comprising one of 350
 CC sequences (ABN79984-ABN80333) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes in particular disease related to homeobox containing
 CC genes (HOX) like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Curarino and limb girdle muscular dystrophy and dwarfism.
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or
 CC (II) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (II) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX Sequence 6089 BP; 1262 A; 219 C; 1585 G; 3023 T; 0 other;
 SQ

Query Match 4.0%; Score 43.4; DB 24; Length 6089;
 Best Local Similarity 45.8%; Pred. No. 0.24;
 Matches 149; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Qy 621 AAAAATGTGTTCTTAATCGAAGAATTTAATCACTGCTTGTGCTAAGAATCAGAAA 680
 Db 1350 AAAAAAATATTAACACCAAAATATATATACACAACTAATAAATCAACAAAA 1291
 Qy 681 GGAGACCGAAGCTTCTAAGTATCAAGAAAAATCTTGGACAATGCTTTCCATAGAG 740
 Db 1290 AACTCTCAACCAATCTCAAAAAATCAAAAAATCTTCTAAAAAATATCTCTAAAT 1231
 Qy 741 AAAACCGTGAGGAAGAGTCGATACCAATAATCCGATTAAAAAGGGAGTCTTACAA 800
 Db 1230 AATATCTTAAAAAATAATTAATTAATTAACAAATAATTTTAAACCAAAAAAACAA 1171
 Qy 801 CAGCCATATCAGTGGAGCAACCTTCCCTTTGATTTCTTTGATTAATGACAGAGATGACAA 860
 Db 1170 AACCACGAAATTCAAAATACCTAAATATTAACACGTTAAAAAATAAATAATATACAAA 1111
 Qy 861 AGAGTAAGTTTTGAATAATCTTAAATAGAAATGCAATTTGTGCGAGCTAAAGCTTGT 920
 Db 1110 AAAAAGCTCAATAAAAAATCTAAAAACACAAACATATTAATAAATAAATAATATACAA 1051
 Qy 921 TCTCTTTATTTCTTTTGTAGAA 945
 Db 1050 TAATCAATTTTATTTTCAAAAAAT 1026

RESULT 4
 AAA91625/C
 ID AAA91625 standard; cDNA; 817 BP.
 AC AAA91625;
 DT 12-JAN-2001 (first entry)
 XX Caenorhabditis elegans cod-5 knockout cDNA.
 DE Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;
 KW Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;
 KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;
 KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes; ss.
 XX Caenorhabditis elegans.
 OS

OS Synthetic.
 XX Key Location/Qualifiers
 FH 40..177 a
 FT CDS /tag= a
 FT /product= "cod-5 frameshift mutant"
 XX
 PN WO200033068-A1.
 XX
 XX 08-JUN-2000.
 XX
 XX 02-DEC-1999; 99WO-US28529.
 XX
 XX 03-DEC-1998; 98US-0205658.
 XX
 XX (GEO) GEN HOSPITAL CORP.
 XX
 XX Ruvkun G, Ogg S;
 XX
 XX WPI; 2000-423022/36.
 XX
 XX Diagnosing and treating obesity and impaired glucose tolerance using
 XX modulators of daf-18 expression and/or activity -
 XX
 XX Disclosure; Fig 43; 402pp; English.
 XX
 XX The present sequence is a Caenorhabditis elegans cod-5 knockout cDNA.
 CC Most of the cod-5 gene has been deleted and this has caused a frameshift
 CC in the remaining coding region. Cod-5 is the aromatic amino acid
 CC hydroxylase that synthesizes serotonin from precursor L-tryptophan. The
 CC cod-5 gene was knocked out to produce mutants completely lacking in
 CC serotonin. The mutants were found to have defects in metabolic control. A
 CC number of C. elegans proteins that have mammalian homologues acting in
 CC the insulin signalling pathway were also identified. The C. elegans age-1
 CC gene encodes a homologue of the mammalian PI 3-kinase whilst daf-2
 CC encodes a homologue of the mammalian insulin receptor. The C. elegans AKT
 CC kinase and PRK kinase act downstream of daf-2 and age-1, just as their
 CC mammalian homologues act downstream of insulin signalling. The C. elegans
 CC PTEN lipid phosphatase homologue, DAF-18, has been found to act upstream
 CC of AKT in the pathway. This discovery has enabled mammalian PTEN action
 CC to be mapped to the insulin signalling pathway. Conserved DAF motifs can
 CC be used to design probes to identify mammalian DAF homologues and thus to
 CC identify individuals with a predisposition towards the development of
 CC glucose intolerance conditions, such as obesity and diabetes.
 XX
 XX Sequence 817 BP; 214 A; 197 C; 166 G; 240 T; 0 other;
 XX
 XX Query Match 3.8%; Score 41.4; DB 21; Length 817;
 XX Best Local Similarity 45.2%; Pred. No. 0.39;
 XX Matches 239; Conservative 0; Mismatches 281; Indels 9; Gaps 2;
 QY 113 TCATCAAAATGCTTATTGAAATTAAGTCTCTGTGGAGTGTATGATTCAAGGGAGA 172
 DB 639 TCCTCAAAATTTCTAGTATGAAATGCTGTGACTGGAAGTAGTAAAGACATCTTGC 580
 QY 173 CGAATATCTGATCCAAATTAAGGAGGACACGATACCTTTATCAATGAAGCGTGCCA 232
 DB 579 TCACACACACATCGGATCAAAACGAATATGTTCCACTACCTCAACGGCATGTGC 520
 QY 233 AGTTCTCTGAGGAGAACTGATGAAGACGGCTCCATATGCTTTCTTCCTTCATGTTTCA 292
 DB 519 AACTCGCCGACACTGCTCAGAAAGTCTGCTCGGTATATCTTTAAATCTTTCATGATTGAT 460
 QY 293 ATAGTCCGCTTTCAACAGTAAACCAAGCAGCGTACAAATAGCGATCAGATTCTTTGT 352
 DB 459 CCAATTTCTTTTACTGGAG-AATCGGAGGGTTCATCAGACGAGACCAAAATCAATCGA 401
 QY 353 AGGTTTGTATGGTGTGTTTTTACTAGGAGAGCTTGTACTTTTCTTCTGACTTTAGTG 412
 DB 400 AAGAAGTAGAGTGTGCAAGCTCTTCAAA-----TCITCTCTGATGCTCCAGA 349
 QY 413 AAGAGTCTCCCATGTTTATGAAATTCAGAAATGAGGGATGAGCAAGCAAGCAG 472

DB 348 GAAGTAATCAATCTCTTGAGAAACTGAGAAAAATCTGATCAGCGAATAGAGCCATG 289
 QY 473 TGGCTAAAGAGTCTATGATGAGATCAGGAGTTTAAAGGAGAAATATATCTTTATCGAGATT 532
 DB 288 TGACCCATGAGCTCGTGAACGGTGTCTGTTCTGGAGTGTAAAAATGATCGCATGATG 229
 QY 533 CGCATTTACAGAGCAATCGGAAGTAAAGGTCCTGCAACAGACAGACATTTGATGGGGA 592
 DB 228 CGAACGTAATTGAGTGCAGAGAGACAGCATATGCAAGACCTGCCAAGAAATCAGAGCT 169
 QY 593 GCTACAAATCTGAAACGGGATATAGGAAAAATGCTGTTCTTAATTCTGA 641
 DB 168 GATAAGTATCCGGCGACTGGCGACACACGGATCCAGTTTTTTCACACTA 120
 RESULT 5
 AAZ17250/c
 ID AAZ17250 standard; cDNA; 868 BP.
 XX
 AC AAZ17250;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:4721.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 DN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 XX (CHIR) CHIRON CORP.
 XX (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 XX WPI; 1999-494092/41.
 XX
 XX Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 XX Claim 1; Page 2243; 2479pp; English.
 XX
 XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene products encoded by one
 CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an

CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX Sequence 868 BP; 148 A; 148 C; 32 G; 284 T; 256 other;

Query Match 3.8%; Score 41; DB 20; Length 868;
 Best Local Similarity 33.1%; Pred. No. 0.51; Indels 0; Gaps 0;
 Matches 89; Conservative 0; Mismatches 180;
 QY 446 AATGAGGAGTGTAGAACCAAGCAGCGTCCCTAAAGGTCATGGATGAGATCAGGAGTT 505
 DB 575 AGCNGNNGANGANNAATTAANNAAGNNGNANGNANANNNTGAGNNAANAGAGGNT 516
 QY 506 AAGGAGAAATTTCTTTATCGAGAGTTCGCTATTACAGAGCAATGGGAAGTAACGGTCC 565
 DB 515 TAGAGNNAANANANANGAGNANNNNNTAGNNAATAATGGNAAATGGANANN 456
 QY 566 TGCAACAGAGACAAGTATTGTGGGGAGCTACAAATCTGAAACGGGATATAGGAAAAA 625
 DB 455 ANTWAGGNANAGNTNNGAGNNGNAGGATGANANATTGANAGAANTNANNNNNNA 396
 QY 626 TGTCTTTCTTAATTCGAGAAATTAATGACTGCTGTGATGGTCTAAGAAATCAGAAAGGAGA 685
 DB 395 NNNATNAATTAATGNNNGNAGGANGNANGAAAAAANTTAAANNANTATNAATAGN 336
 QY 686 CCGAAAGCTTCTAAGTAATCAAGAAAAA 714
 DB 335 NNGAAGNTNANGNATGNTANANNA 307

RESULT 6
 ID ABK28408 standard; DNA; 9504 BP.

XX AC ABK28408;
 DT 23-APR-2002 (first entry)
 XX DE DNA transcription associated complementary genomic DNA #141.
 XX KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.

XX OS Unidentified.

XX PN WO200192565-A2.

XX PD 06-DEC-2001.

XX PF 06-APR-2001; 2001WO-BP03973.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-090046/12.
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 PT tumours or cancer.
 XX Claim 1; SEQ ID No 282; 32pp; English.
 PS The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculous, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 9504 BP; 2775 A; 138 C; 2348 G; 4243 T; 0 other;

Query Match 3.8%; Score 41; DB 24; Length 9504;
 Best Local Similarity 53.4%; Pred. No. 1.2;
 Matches 86; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 862 GAGTAAGTTTTTTGAAAAATCTAAATAGAAATGCGATTTGTGCGAGCTAAAGCTTGGTT 921
 DB 8911 GTGGATGATATATATTTGGTGTATTTTGGTTTATTGCTATGGGGCTATTTGTTGTTT 8970
 QY 922 CTCCTTTATTTTCCTTTTGTAGAAATGATTCGGTAGTAGGAATATGGGTGCGAGTATGCACG 981
 DB 8971 ATTGTTATTGTTGTTTTTTTATGATTTTTTTTAAAGAAAGTGGGTTAGGTTTAGAG 9030
 QY 982 TTGGAGTGTATGCTGCTCTCGAAGAACAGGAGGAAAGTGATT 1022
 DB 9031 ACATAGTGTGTTTTTTGTTTAGGGGAATAGGATATTAATT 9071

RESULT 7

ID ABJ92307 standard; DNA; 5152 BP.

XX AC ABJ92307;

XX DT 01-JUL-2002 (first entry)

XX DE Chemically treated DNA repair gene fragment complementary to#58.

XX KW DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3;
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4;
 KW DDITL; FANCB; XRCC8; ataxia telangiectasia; aging; Bloom's syndrome;
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;
 KW cancer; ds.

XX OS Unidentified.

XX PN WO200181622-A2.

XX PD 01-NOV-2001.

XX 06-APR-2001; 2001WO-EP03972.
 XX PF
 XX 06-APR-2000; 2000DE-1019058.
 XX PR
 XX 07-APR-2000; 2000DE-1019173.
 XX PR
 XX 30-JUN-2000; 2000DE-1032529.
 XX PR
 XX 01-SEP-2000; 2000DE-1043826.
 XX PR
 XX (EPIG-) EPIGENOMICS AG.
 XX PA
 XX Olek A, Piepenbrock C, Berlin K;
 XX PI
 XX WPI; 2002-034446/04.
 XX DR
 XX New nucleic acid derived from genes associated with DNA repair, useful
 XX PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
 XX PT cytosine methylation -
 XX PS
 XX Claim 1; SEQ ID NO 116; 25pp + sequence listing; English.
 XX CC The invention relates to nucleic acids containing a sequence of at least
 XX CC 18 nucleotides of chemically treated DNA of genes associated with DNA
 XX CC repair and their complements. The invention also relates to nucleic
 XX CC acids comprising at least 18 base pairs of the chemically pretreated DNA
 XX CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
 XX CC PMS2L2, PMS2L5, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDP1, TDG, INPPL1,
 XX CC RFC4, DDIRL, FANCB, or ERCC8. Nucleic acids of the invention and related
 XX CC oligomers, are useful for diagnosis of diseases associated with gene
 XX CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
 XX CC Cockayne syndrome, Ni-jmegen breakage syndrome or Werner syndrome.
 XX CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours
 XX CC and cancer, particularly by determining status of cytosine methylation
 XX CC and/or by detecting single-nucleotide polymorphisms. Determination of
 XX CC individual methylation patterns may allow development of individualised
 XX CC therapies. The sequences given in records ABL92192-ABL92335 represent
 XX CC chemically pre-treated DNA fragments from genes associated with DNA
 XX CC repair, and their complements.
 XX CC Note: The sequence data for this patent is not represented in the
 XX CC specification, but is based on sequence information supplied by the
 XX CC European Patent Office.
 XX SQ
 XX Sequence 5152 BP; 1608 A; 36 C; 965 G; 2543 T; 0 other;
 Query Match 3.7%; Score 40.6; DB 24; Length 5152;
 Best Local Similarity 44.6%; Pred. No. 1.3;
 Matches 201; Conservative 0; Mismatches 249; Indels 1; Gaps 1;
 OY 584 TGATGGGAGCTACAAATCTGAAACGGGATATAGGAAATCTGTTCTTAATTCGAAG 643
 DB 73 TGTGGGGTGGGGGATATATTTATTTGAGTTTGAAGTTTGAATTTGAGTGGCTCGA 132
 OY 644 AATTTAATGACTGCTGATGCTTGAATATCGAATATCGAAGAGACCGAAGCTTCTAGTAA 703
 DB 133 GTATTTTATGTTGGGTATGATGTCGATTTTGTATATAAATAGATTTAGTTAGTATAA 192
 OY 704 TCAAGAAAAATCTTTGGACAATAGCTTTTCCATPAGAGAAACCGGTGAGGAAGAGTCGA 763
 DB 193 GATTTTAAAAATTTTAAATATTTTTTTTTTTTTTTTGGGAAAGTGGAATAAAGGAACGG 252
 OY 764 TACCATTAATTCGATTTAAAAAGGGAGTGTCTTACACAGCCATATCATGTGGGACCACT 823
 DB 253 ATTTTATTTTTTTTTTTTGGAGGAATAAATTTTGAATATAGATATTTAATGTAGATTTG 312
 OY 824 TCCTTTGATCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 883
 DB 313 AGAATGCTGATCTTTAT-TTTAAGTAGCTGTTTATATATATATATATTTTGAATAATTT 371
 OY 884 AAAATAGAAATCCATCTGTCGAGCTTAAGCTTCTCTTCTTCTTCTTCTTCTTCTTCTG 943
 DB 372 TATTTAATATTTATTTATTTTATGTTTATTTATTTATTTATTTATTTATTTATTTATTT 431
 OY 944 ATGATTCGCTAGTAAATATGCGGTGCGAGTATGACGCTTGGAGTATGCTCTTCTTGG 1003

DB 432 GTGTTTTACGAAAAAATAATAATGTTTTTTTATTTTTTAAAGTGTATATAATTTTTTTAGT 491
 OY 1004 AAGAACAGGAAAGTATGCTGCTCAGCTTTA 1034
 DB 492 TAATATTTCTGTAAGAAATTTTTTAGATTTTA 522
 RESULT 8
 ABL49374
 ID ABL49374 standard; DNA; 5152 BP.
 XX AC
 XX ABL49374;
 XX DT
 XX 01-MAY-2002 (first entry)
 XX DE Human polynucleotide associated with DNA replication SEQ ID NO 74.
 XX KW Human; cytostatic; neuroprotective; nootropic; immunostimulant;
 XX KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
 XX KW ERCC3; SNRPAL; RAD50; LIG2; cytosine methylation; Ataxia telangiectasia;
 XX KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
 XX OS Homo sapiens.
 XX WO200177377-A2.
 XX PD 18-OCT-2001.
 XX PF 06-APR-2001; 2001WO-EP03971.
 XX PR 06-APR-2000; 2000DE-1019058.
 XX PR 07-APR-2000; 2000DE-1019173.
 XX PR 30-JUN-2000; 2000DE-1032529.
 XX PR 01-SEP-2000; 2000DE-1043826.
 XX PA (EPIG-) EPIGENOMICS AG.
 XX PI Olek A, Piepenbrock C, Berlin K;
 XX DR WPI; 2002-017471/02.
 XX PT New nucleic acid sequences from chemically modified genes associated
 XX PT with DNA replication, useful for analysing cytosine methylations for
 XX PS diagnosis and therapy of diseases e.g. Ataxia telangiectasia -
 XX PS Claim 1; SEQ ID NO 74; 23pp + Sequence Listing; English.
 XX CC The invention relates to nucleic acid sequences comprising at least 18
 XX CC bases of a chemically pretreated gene associated with gene regulation,
 XX CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences
 XX CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPAL,
 XX CC RAD50 and LIG2. The chemical pretreatment converts cytosine bases
 XX CC unmethylated at the 5-position to uracil or another base with
 XX CC hybridisation behaviour dissimilar to cytosine, to enable analysis of
 XX CC cytosine methylations. The DNA sequences and method are useful in the
 XX CC diagnosis of diseases (or predisposition to diseases) associated with DNA
 XX CC replication and in therapy of such diseases, by enabling analysis of the
 XX CC cytosine methylation patterns of such genes. They are especially useful
 XX CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's
 XX CC syndrome, solid tumours and cancer.
 XX CC Note: The sequence data for this patent did not form part of the printed
 XX CC specification and was supposed to be available directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences. However, the sequence data did
 XX CC not correspond to that referred to in the specification. The present data
 XX CC is taken from EPO data for the patent.
 XX SQ
 XX Sequence 5152 BP; 1608 A; 36 C; 965 G; 2543 T; 0 other;
 Query Match 3.7%; Score 40.6; DB 24; Length 5152;
 Best Local Similarity 44.6%; Pred. No. 1.3;
 Matches 201; Conservative 0; Mismatches 249; Indels 1; Gaps 1;
 OY 584 TGATGGGAGCTACAAATCTGAAACGGGATATAGGAAATGTTTCTTAATTCGAAG 643

Db 73 TGTGGGTGGGAGGATATTTATTTAGTTAGTAATTTGAAATTTAGTGTAGTCTGA 132
Qy 644 AATTTAATGACTGCTGATGGTCTAAGAAATCAGAAAGGAGACCGAAGCTTCTAAGTAA 703
Db 133 GTATTTAGTTGGTGAATAGTGTGAGATTTTGTATATAAATAAGATTTAGTTAGTATA 192
Qy 704 TCAAGAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAAACCGTAGGAAAGGAGTGA 763
Db 193 GATTTTAAAAATTTTAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 252
Qy 764 TACCATAAATTCGATTAATAAAGGAGTGTACACACGCCATATCAGTCGAGCAACCT 823
Db 253 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 312
Qy 824 TCCCTTTGATCTTTGATATCAGACAGATGACAAAGAGTAAGTTTGTGAAATTTCT 883
Db 313 AGAATGTTGATGTTTATTTTAAAGTAGGTGTTTATATATATATTTATAGAAATTTT 371
Qy 884 AATAAGAAATGCAATTTGTCGAGCTAAGCTTCTCTTTATTTTCCCTTTCTGAGA 943
Db 372 TATTTTAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 431
Qy 944 ATGATTCGAGTAGTAAGAAATAGGGTCGAGATGACACCTTGGAGTGAATGGCTGTTCTGG 1003
Db 432 CTGTTTATACGAAAAAATAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 491
Qy 1004 AAGAACAGGGAAGTGAATTTCTCAGCTTTA 1034
Db 492 TAATATTTCTGAAGAAATTTTATAGATTTA 522

RESULT 9
ABQ67077
ID ABQ67077 standard; DNA; 37184 BP.
XX AC ABQ67077;
XX DT 28-AUG-2002 (first entry)
XX DE Human angio genesis associated polynucleotide SEQ ID NO 107.
XX KW Human; angio genesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX OS Homo sapiens.
XX PN WO200246454-A2.
XX PD 13-JUN-2002.
XX PF 06-DEC-2001; 2001WO-EP14320.
XX PR 06-DEC-2000; 2000DE-1061338.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Schacht O;
XX WPI; 2002-500450/53.
XX New nucleic acid fragments from chemically treated
PT angio genesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer
XX
PS Claim 1; SEQ ID NO 107; 41pp + Sequence Listing; German.
XX The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angio genesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also

CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angio genesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 37184 BP; 8680 A; 717 C; 9540 G; 18247 T; 0 other;

Query Match 3.7%; Score 40.6; DB 24; Length 37184;
Best Local Similarity 57.5%; Pred. No. 2.7;
Matches 73; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
Qy 834 TCTTTGATAATGACAGAGATGACAAAGAGTAAGTTTGTGAAAAATTTCTAAATAGAAA 893
Db 6242 TTTTAAAGATAGATGGATGTAGAAATAGGTGTATGTATTAGAGTTTGTATAAAAAAG 6301
Qy 894 TGCATTTGTGCGAGCTAAAGCTTCTCTTTATTTTCCCTTTGTAGAAATGATTCGGT 953
Db 6302 TGTATTTGTGTTAGGTATAGTGGTTTATGTATTGTAAATTTTAGTATTTCGGAGGTTGAGGT 6361
Qy 954 ACTAGGA 960
Db 6362 AGGAGGA 6368

RESULT 10
ABA92787
ID ABA92787 standard; DNA; 640681 BP.
XX AC ABA92787;
XX DT 27-MAR-2002 (first entry)
XX DE Buchnera sp. genomic DNA SEQ ID NO:1.
XX KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
XX OS Buchnera sp.
XX PN JP2001292771-A.
XX PD 23-OCT-2001.
XX PF 07-APR-2000; 2000JP-0107160.
XX PR 07-APR-2000; 2000JP-0107160.
XX PA (RIKA) RIKAGAKU KENKYUSHO.
XX DR WPI; 2002-126043/17.
XX PT A genomic DNA of cockroach-symbiotic bacterium -
XX PS Claim 1; Page 16-230; 237pp; Japanese.
XX

XX The present invention describes a gene (I) derived from Buchnera sp.
CC containing the DNA (a) or (b), (a) has a fully defined base pair
CC sequence selected from a table of sequences found in the Buchnera sp.
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (i); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC (d) (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and

DR WPI; 2001-602751/68.
 XX Designing primers and probes for analysing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle -
 XX
 XX Claim 1; SEQ ID NO 98; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.
 XX
 XX Sequence 17419 BP; 5015 A; 311 C; 3864 G; 8229 T; 0 other;
 SQ

Query Match 3.6%; Score 39.6; DB 22; Length 17419;
 Best Local Similarity 49.5%; Pred. No. 3.7;
 Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 852 GATCAACAAGAGTAAGTCTTTCGAAAATTCCTAAATAGAAATCCATTGTCGAGCTA 911
 DB 2146 GTTTAGTAATATTAGAAATGTTGAATGGTATAGAAAGATATATTTAGATTTTA 2205
 QY 912 AAGCTTCCTCTCTTTATTTCTTTTGTAGAAATCGTAGAGGATATGCGGTCG 971
 DB 2206 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2265
 QY 972 AGTATGACCTTCGAGTATGCTGCTCTCGAAGACAGGAAAGTATGTCACGCT 1031
 DB 2266 ATTATTTAGTGGTGGAGTAGTGATGATATATAGTTTATTTATTTTATTTTGGG 2325
 QY 1032 TTAGAGCAATCTTCAGATATACCTTT 1057
 DB 2326 TTAAAGTATTTTATTTTATTTAGTTT 2351

RESULT 13
 ABL33295
 ID ABL33295 standard; DNA; 17419 BP.
 XX
 XX ABL33295;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Human immune system associated gene SEQ ID NO: 1268.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytostatic; neotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritis; antidabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EF07537.
 PF

XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PT
 XX WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 XX Claim 1; SEQ ID NO 1268; 32pp + Sequence Listing; German.
 PS
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 XX Sequence 17419 BP; 5015 A; 311 C; 3864 G; 8229 T; 0 other;
 SQ

Query Match 3.6%; Score 39.6; DB 24; Length 17419;
 Best Local Similarity 49.5%; Pred. No. 3.7;
 Matches 102; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 852 GATCAACAAGAGTAAGTCTTTCGAAAATTCCTAAATAGAAATCGATTGTCGAGCTA 911
 DB 2146 GTTTAGTAATATTAGAAATGTTGAATGGTATAGAAAGATATATTTAGATTTTA 2205
 QY 912 AAGCTTCCTCTCTTTATTTCTTTTGTAGAAATCGTAGAGGATATGCGGTCG 971
 DB 2206 AATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2265
 QY 972 AGTATGACCTTCGAGTATGCTGCTCTCGAAGACAGGAAAGTATGTCACGCT 1031
 DB 2266 ATTATTTAGTGGTGGAGTAGTGATGATATATAGTTTATTTATTTTATTTTGGG 2325
 QY 1032 TTAGAGCAATCTTCAGATATACCTTT 1057
 DB 2326 TTAAAGTATTTTATTTTATTTAGTTT 2351

RESULT 14
 ABK28238
 ID ABK28238 standard; DNA; 17419 BP.
 XX
 XX ABK28238;
 XX
 XX 23-APR-2002 (first entry)
 DT
 XX DNA transcription associated complementary genomic DNA #56.
 DE
 XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Kieffer's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.
 XX
 XX Unidentified.
 OS
 XX WO200192565-A2.
 PN
 XX

XX	DT	04-APR-2001 (first entry)	
XX	DE	Genomic fragment #37.	
XX	XX	Genomic library; bacteria; human upper airway; otitis media; sinusitis;	
XX	KW	bronchopulmonary; endocarditis; meningitis; ss.	
XX	XX	Moraxella catarrhalis.	
XX	OS		
XX	PN	WO200078968-A2.	
XX	PD	28-DEC-2000.	
XX	PF	16-JUN-2000; 2000WO-US16649.	
XX	PR	18-JUN-1999; 99US-0140121.	
XX	XX	(INCY-) INCYTE GENOMICS INC.	
XX	PA		
XX	PI	Lagace RE, Patterson C, Berg KL;	
XX	XX	WPI; 2001-041427/05.	
XX	XX	Genomic library for identifying diagnostic and therapeutic	
XX	PT	compositions, and for identifying virulence factors, regulatory	
XX	PT	elements and drug targets, comprises Moraxella catarrhalis nucleic	
XX	PT	acids	
XX	XX	Claim 1; Page 391-415; 545pp; English.	
XX	PS	The present invention relates to a Moraxella catarrhalis genomic library	
XX	CC	comprising of a combination of 41 nucleic acid molecules (see	
XX	CC	AAE28514-2AF28554). The library has a number of uses described in the	
XX	CC	specification. e.g. is useful for identifying diagnostic and therapeutic	
XX	CC	compositions. M. catarrhalis (Branhamella catarrhalis) is a large	
XX	CC	aerobic gram-negative diplococcus, normally found among the bacterial	
XX	CC	flora of human upper airways. M. catarrhalis is known to cause acute,	
XX	CC	localised infections such as otitis media, sinusitis and bronchopulmonary	
XX	CC	infection and life-threatening, systemic diseases including endocarditis	
XX	CC	and meningitis.	
XX	XX	Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;	
XX	SQ		
XX	XX	Query Match 3.6%; Score 39.6; DB 22; Length 99629;	
XX	XX	Best Local Similarity 61.8%; Pred. No. 7.2;	
XX	XX	Matches 63; Conservative 0; Mismatches 39; Indels 0; Gaps 0	
XX	QY	803 GCATATCATGTGGAGCAACCTTCCTTTGGATCTTTTCATATGACAGAGATGACACG 862	
XX	Db	8570 GCATATTCGCGATGCAATTTACATTTAAATATTTCAAACCTGCAATTTGACCAAG 8511	
XX	QY	863 AGTAAGTTTTGAAAAATTTCTAAAATAGAAATGCAATTTGTGT 904	
XX	Db	8510 ATGAATTTTTGAAAAATGCCAACGGCAATACTATGTGT 8469	
XX	XX	RESULT 16	
XX	ID	ABL32333/C	
XX	XX	ABL32333 standard; DNA; 8895 BP.	
XX	AC	ABL32333;	
XX	XX	26-MAR-2002 (first entry)	
XX	DT	Human immune system associated gene SEQ ID NO: 306.	
XX	DE	Human; immune system disease; cytosine methylation; antiasthmatic;	
XX	KW	antiarteriosclerotic; antianemic; cytostatic; neutrophic;	
XX	KW	antiprotective; anti-HIV; anticonvulsant; ophthalmological;	
XX	KW	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;	
XX	KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anacmia;	
XX	KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
OS Homo sapiens.
PN WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP07537.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX (EPIC-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX Claim 1; SEQ ID NO 306; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX Sequence 8895 BP; 2269 A; 62 C; 1898 G; 4665 T; 1 other;
SQ
Query Match 3.6%; Score 39; DB 24; Length 8895;
Best Local Similarity 46.2%; Pred. No. 4.2;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
Qy 635 AATTGGAAGAAATTTAATGACTGCTTGATGGTCTAAGAAATCAGAAAGGAGCCGAAAGCT 694
Db 7126 AATACAAAAAATAAACCACCTAAATAAAAAAATATACATAAAAAAATAA 7067
Qy 695 TCTAAGTAATCAAGAAAAAATCTGGACAATAGCTTTTCCATAGAGAAACCGTGAGGAA 754
Db 7066 AAAAAAACTAAAAAATAAATTAACAAACAAAAAATAATATATAAAAACCAAA 7007
Qy 755 AGGAGTCGATACCAATAATTCGATTAAAAAGGGAGTGCTTACACACGCATATACGTG 814
Db 7006 AAAAAAATAAATAATATCGAAAAACCTTAATACTACTACCTAAACCTCTTACAT 6947
Qy 815 GAGCAACCTTCCCTTTGATCTTTGATTAATGACAGAGATGAACAAGAGTAAGTTTGG 874
Db 6946 CTCGCCAAACGACCTTTTACCCATCTATAAAAACAATAACCAATCTAATTAATATCTAA 6887
Qy 875 AAAAATCTAAATAAGAAATGCAATTTGTGTCGAGCTAAA 913
Db 6886 AAATAACCTAAAAATCAAAACCAACCTCTTTCGTGATACA 6848
RESULT 17
ABV61126/c
ID ABV61126 standard; cDNA; 326 BP.
XX
XX ABV61126;
XX
XX 13-SEP-2002 (first entry)
XX Human prostate expression marker cDNA 61117.
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
PN WO200160860-A2.
XX 23-AUG-2001.
XX 20-FEB-2001; 2001WO-US05171.
XX 17-FEB-2000; 2000US-183319P.
XX 16-MAR-2000; 2000US-189862P.
XX 25-MAY-2000; 2000US-207454P.
XX 09-JUN-2000; 2000US-211314P.
XX 18-JUL-2000; 2000US-219007P.
XX 13-DEC-2000; 2000US-255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
XX Claim 1; Page 11610; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 326 BP; 100 A; 48 C; 37 G; 140 T; 1 other;
SQ
Query Match 3.6%; Score 38.6; DB 23; Length 326;
Best Local Similarity 49.0%; Pred. No. 1.6; 105; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
Qy 735 ATAGAGAAACCGTGAGGAAAGAGTCGATACCATATAATTCGGATTAAAAAGGAGTCG 794
Db 268 AAAACCAACCAACCCGGGGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 209
Qy 795 TTACAACAGCCATATCAGTCGAGCAACCTTCCCTTTGATTTCTTTGATAATGACAGAAGAT 854
Db 208 TTACAAAAGGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 149
Qy 855 GAACAAGAGTAAGTTTGTGAAAAATTTCTAAAAATAGAAATGCATTTTGTGTCGAGCTAAAG 914
Db 148 AACAAAAAATAAATTTTGTAAAAATAAATAAATAAATAAATAAATAAATAAATAA 89
Qy 915 CTGCTTCTCTTTTATTTTCTTTTGT 940
Db 88 TTTTCTTTTCTTTTCTTTTCTTTT 63
RESULT 18
ABK33948
ID ABK33948 standard; DNA; 18997 BP.
XX
XX ABK33948;

18-JUN-2002 (first entry)

Human DNA for staging of Astrocytomas #16.

Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG; bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry; matrix assisted laser desorption/ionization mass spectrometry.

Homo sapiens.

W0200202808-A2.

10-JAN-2002.

02-JUL-2001; 2001WO-EP07538.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI: 2002-171649/22.

Novel chemically modified genomic DNA sequences, useful in the characterisation, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas or predisposition to astrocytomas -

claim 1; SEQ ID No 31; 37pp; English.

The invention relates to a nucleic acid comprising a sequence (I) of at least 18 bases in length of a segment of chemically pre-treated genomic DNA which has any one of the sequences of (ABK33919-ABK34032) or its complement. Also included are an oligonucleotide or peptide nucleic acid (or set thereof) of at least 9 nucleotides which hybridises to (I), primers for (I), probes for detecting cytosine methylation or single-nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide nucleic acids for analysing diseases associated with the methylation states of the CpG dinucleotides of (I). The array is useful for determining genetic and/or epigenetic parameters, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas, or the predisposition to astrocytomas by analysing genomic methylations, involves obtaining a biological sample containing cytosine DNA, extracting the genomic DNA, converting cytosine bases, which are unmethylated at the 5-position, in the genomic DNA sample, to uracil or another base which is dissimilar to cytosine in terms of hybridisation behaviour, by chemical treatment and amplifying chemically pre-treated genomic DNA fragments using the array and a polymerase, where the genomic DNA fragments carry a detectable label. The method further involves identifying methylation status of one or more cytosine positions, and analysing methylation status of the cytosine positions by reference to one or more data sets, the genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplificates carry a fluorescent label or radionuclide. Optionally, the labels of the amplificates are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplificates or fragments of the amplificates are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published.pct.sequences.

XX Sequence 18997 BP; 4675 A; 299 C; 4016 G; 10007 T; 0 other;
SQ

Query Match 3.5%; Score 38.4; DB 24; Length 18997;
Best local similarity 54.2%; Pred. Nish. 8.2;
Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps

QY 856 ACAAGAGGTAAGTCTTTGAATAATTCATAAATAGAATGCATTGTGTCGAGCTAAAGC 915
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4017 AAAAAGAAAAAAAANATAGTAAATTTTAAATATTTTATTTGTTAAATATAAAAT 4076

QY 916 TTGGTCTCTCTATTATTTCCCTTTTGAGAAATGATTCGGTAGTACGATATGGCGTCGAGTA 975
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4077 TTGAATAATTAATTTAGTTTTTTTTTTTGAATAATGAGTTGGAGAGATAGGGTTGANTTAGTA 4136

QY 976 TGCACGTTGGAGCTGATTTGGCTGTT 999
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4137 TGTTTTTTCGGGTTTTTCGACGTT 4160

RESULT 19
ABL32570
ID ABL32570 standard; DNA; 18997 BP.
XX XX
AC AC
ABL32570;
XX XX
DT DT
26-MAR-2002 (first entry)
XX XX
DE DE
Human immune system associated gene SEQ ID NO: 543.
XX XX
Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiatherosclerotic; anti-tubercular; cytostatic; neurotropic;
KW neuroprotective; anti-HIV; anticongestant; ophthalmological;
KW antirheumatic; antarthritic; antidiabetic; antipsoriatic;
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX XX
OS Homo sapiens.
XX XX
WO200200928-A2.
XX XX
ED 03-JAN-2002.
XX XX
EF 02-JUL-2001; 2001WO-EP07537.
XX XX
PR 30-JUN-2000; 2000DE-1032529.
XX XX
PR 01-SEP-2000; 2000DE-1043826.
XX XX
PA (EPIG-) EPIGENOMICS AG.
XX XX
PI Olek A, Piepenbrock C, Berlin K;
XX XX
DR WPI; 2002-130909/17.
XX XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX XX
PS Claim 1; SEQ ID NO 543; 32pp + Sequence Listing; German.
XX XX
The present invention provides a number of human immune system associ-
CC genes which are modified by the methylation of cytosines. The sequence
CC can be used in the diagnosis and treatment of immune system disorders
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloi-
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/alcalterative bowel
CC diseases. The present sequence is a gene of the invention.
XX XX
SO Sequence 18997 BP; 4675 A; 299 C; 4016 G; 10007 T; 0 other;

Query Match 3.5%; Score 38.4; DB 24; Length 18997;
 Best Local Similarity 54.2%; Pred. No. 8.2;
 Matches 78; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 856 AACAAAGAGTAAGTCTTTTCAAAAATCTAAATAGAAATGCAATTTGTCGAGCTAAAGC 915
 DB 4017 AAAAAAAAAAAAAAAGTCTAAATTTTAAATAATTTTCTTAAATATATAAAT 4076

QY 916 TTGCTCTCTTATTTCTTTTCTTTAGATATCGTCTAGTATAGTATGCGGTCGAGTA 975
 DB 4077 TTGAATATTTAGTTTTTTTTTTTAAATAGTCTGGAGAGTAGGTTGAATAGTA 4136

QY 976 TGCACGTTGAGTGATGCTGCTTT 999
 DB 4137 TGTTTTTCGGGTTTTTCACGCTT 4160

RESULT 20
 AAD26400/c
 ID AAD26400 standard; DNA; 50000 BP.
 XX AAD26400;
 AC
 DT
 DT 26-MAR-2002 (first entry)
 XX Human glutamate receptor, metabotropic 3 (GRM3) gene fragment #1.
 DE Human;
 KW Human; glutamate receptor metabotropic 3; GRM3; neuroprotective;
 KW nootropic; gene therapy; neurotransmitter; chromosome 7q21.1-q21.2;
 KW drug screening; neurological disorder; polymorphism; ds.
 XX Homo sapiens.

XX Key Location/Qualifiers
 FT variation replace (4006, A)
 FT /tag= a
 FT /note= "Nucleotide R is present at this location in the
 FT sequence shown in sequence listing of the specification"
 FT 4067..4528
 FT /tag= b
 FT /number= 1
 FT variation replace (4192, A)
 FT /tag= c
 FT /note= "Nucleotide R is present at this location in the
 FT sequence shown in sequence listing of the specification"
 FT 4529..25175
 FT /tag= d
 FT /number= 25176..26031
 FT /tag= e
 FT variation replace (25586, T)
 FT /tag= f
 FT /note= "Nucleotide Y is present at this location in the
 FT sequence shown in sequence listing of the specification"
 FT 26032..50000
 FT /tag= g
 FT /partial
 FT variation replace (26157, A)
 FT /tag= h
 FT /note= "Nucleotide M is present at this location in the
 FT sequence shown in sequence listing of the specification"
 XX PN W0200196350-A2.
 XX
 XX 20-DEC-2001.
 XX
 XX 18-JUN-2001; 2001WO-US19447.
 XX
 XX 16-JUN-2000; 2000US-212328P.
 XX
 XX (GENA-) GENAISSANCE PHARM INC.
 XX
 XX Finkel K, Koshy B, Tanguay DA;
 PI

XX WPI: 2002-090198/12.
 XX New isolated polynucleotide, a polymorphic variant of glutamate
 PT receptor, metabotropic 3 (GRM3) gene for expressing GRM3 protein
 PT isoform to screen drugs to treat GRM3 activity-related disease -
 XX
 XX Example 1; Fig 1; 165pp; English.
 XX
 CC The invention relates to an isolated polynucleotide which is a
 CC polymorphic variant of glutamate receptor, metabotropic 3 (GRM3)
 CC isogene. GRM3 is a receptor for glutamate, the major excitatory
 CC neurotransmitter in the mammalian central nervous system. Human GRM3
 CC located on chromosome 7q21.1-q21.2 is expressed in human foetal and adult
 CC whole brain especially in the caudate nucleus and corpus collosum. GRM3
 CC DNA is useful in gene therapy and also for studying the expression and
 CC function of GRM3. GRM3 polypeptide is used for screening drugs. A
 CC recombinant non-human organism is used to study expression of GRM3 SG
 CC in vivo, for in vivo screening and testing of drugs targetted against
 CC GRM3 protein, and for testing the efficacy of therapeutic agents and
 CC compounds for neurological disorders in a biological system. GRM3
 CC haplotypes are for treating diseases associated with GRM3 activity,
 CC e.g., neurological disorders. The present sequence is human GRM3 gene
 CC fragment containing polymorphic sites.
 XX
 XX Sequence 50000 BP; 15337 A; 9556 C; 9289 G; 15817 T; 1 other;
 QY 727 GCTTTTCCATAGAGAAACCGTGAGGAGGAGTCGATACCATTAATTCGATTAATAAA 786
 DB 41412 GCCTTTCTCTATGATGTCATTCTTTATTAAGACATCATTAACAATATACATTATAACAAC 41353

QY 787 GGGAGTGCTTACACAGCCATATCATGTCGAGCAACCTTCCCTTTTGTGATTCATGTAATGA 846
 DB 41352 TAATGAGTTCACAAATACCTCTCTTATTTATTCAGATTTGGAAACAATATTTATTATTA 41293

QY 847 CAGAGAGTGAACAAAGAGTAAGTTTGTGAAAAATTTCAAAATAGAAATGCAATTTGTGTGCG 906
 DB 41292 CTGTGAATGCACATAG-GTAAGCTGAGCCAAATGACATGAGACTGCATCTTTTCTTAA 41234

QY 907 AGCTAAAGCTTGCTTCCTCTCTTTTATTTTCCTTTT 938
 DB 41233 ATACAAACCTTGTTTTTTTTTCTTTCTTTT 41202

RESULT 21
 AAD26437/c
 ID AAD26437 standard; DNA; 50000 BP.
 XX AAD26437;
 AC
 DT 26-MAR-2002 (first entry)
 XX Human GRM3 gene fragment #1 allelic variant containing polymorphic site.
 DE Human;
 KW Human; glutamate receptor metabotropic 3; GRM3; neuroprotective;
 KW nootropic; gene therapy; neurotransmitter; chromosome 7q21.1-q21.2;
 KW drug screening; neurological disorder; polymorphic site; PS; ds.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT misc_feature. 4006
 FT /tag= a
 FT /note= "This degenerate base represents polymorphic site
 FT (PS) 1"
 FT misc_feature 4192
 FT /tag= b
 FT /note= "This degenerate base represents polymorphic site
 FT (PS) 2"
 FT

```
FT misc_feature 25586 /*tag= c
FT FT /note="This degenerate base represents polymorphic site
FT (PS) 3"
FT misc_feature 26157 /*tag= d
FT FT /note="This degenerate base represents polymorphic site
FT (PS) 4"
XX WO200196350-A2.
XX
XX 20-DEC-2001.
XX
XX 18-JUN-2001; 2001WO-US19447.
XX
XX 16-JUN-2000; 2000US-212328P.
XX
XX (GENA-) GENAISSANCE PHARM INC.
XX
XX Finkel K, Koshy B, Tanguay DA;
XX
XX WPI; 2002-090198/12.
XX
XX
XX New isolated polynucleotide, a polymorphic variant of glutamate
XX receptor, metabotropic 3 (GRM3) gene for expressing GRM3 protein
XX isoform to screen drugs to treat GRM3 activity-related disease -
XX
XX Claim 20; Page 135-153; 165pp; English.
XX
XX The invention relates to an isolated polynucleotide which is a
XX polymorphic variant of glutamate receptor, metabotropic 3 (GRM3)
XX isogene. GRM3 is a receptor for glutamate, the major excitatory
XX neurotransmitter in the mammalian central nervous system. Human GRM3
XX located on chromosome 7q21.1-q21.2 is expressed in human foetal and adult
XX whole brain especially in the caudate nucleus and corpus collosum. GRM3
XX DNA is useful in gene therapy and also for studying the expression and
XX function of GRM3. GRM3 polypeptide is used for screening drugs. A
XX recombinant non-human organism is used to study expression of GRM3 SG
XX in vivo, for in vivo screening and testing of drugs targeted against
XX GRM3 protein, and for testing the efficacy of therapeutic agents and
XX compounds for neurological disorders in a biological system. GRM3
XX haplotypes are for treating diseases associated with GRM3 activity,
XX e.g., neurological disorders. The present sequence is an allelic variant
XX of human GRM3 gene fragment containing polymorphic sites (PS).
XX
XX Sequence 50000 BP; 15337 A; 9554 C; 9287 G; 15818 T; 4 other;
XX
XX Query Match 3.5%; Score 38.4; DB 24; Length 50000;
XX Best Local Similarity 51.9%; Pred. No. 12;
XX Matches 110; Conservative 0; Mismatches 101; Indels 1; Gaps 1;
XX
XX QY 727 GCTTTCCATACAGAAACCGTGGAGAAAGGAGTCCGATACCATATAATCCGATTAAAAA 786
XX DB 41412 GCCTTCCATACAGTATCTTTTAAAGACATATTATACATTATACAAAC 41353
XX
XX QY 787 GGGAGTGCCTACAGACCATATCAGTGGAGCAACCTCCCTTTCATCTTGAATATGA 846
XX DB 41352 TAATGAGTTACAAATACCTCTTATATTATTCAGATTGGAAACAATTTTATTATTA 41293
XX
XX QY 847 CAGAAATGACAAAGTAAAGTATTTTGAAGAAATTCCTAAATAGAAATGCAATTCGTCG 906
XX DB 41292 CTGTGAATGCATAG-GTAAAGCTGAGCCAAATGACAAATGACATGCTATCTTTCTAA 41234
XX
XX QY 907 AGCTAAAGCTGCTCTCTCTTTATTTCTTTT 938
XX DB 41233 ATACAAACCTTGTTTTTTTCTTTCTTTT 41202
XX
XX RESULT 22
XX ABQ67094
XX ID ABQ67094 standard; DNA; 83391 BP.
XX
XX ABQ67094;
```

```
XX
XX 28-AUG-2002 (first entry)
XX
XX Human angiogenesis associated polynucleotide SEQ ID NO 124.
XX
XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
XX inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
XX macular degeneration; inflammatory bowel disease; Crohn's disease;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiarteriosclerotic; ds.
XX
XX Homo sapiens.
XX
XX WO200246454-A2.
XX
XX 13-JUN-2002.
XX
XX 06-DEC-2001; 2001WO-EP14320.
XX
XX 06-DEC-2000; 2000DE-1061338.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Schacht O;
XX
XX WPI; 2002-500450/53.
XX
XX New nucleic acid fragments from chemically treated
XX angiogenesis-associated genes, useful for determining methylation
XX status, e.g. in diagnosis or treatment of cancer -
XX
XX Claim 1; SEQ ID NO 124; 41pp + Sequence Listing; German.
XX
XX The invention relates to a nucleic acid (I) comprising a segment of 18
XX bases of chemically pretreated DNA of angiogenesis-associated genes (II)
XX having sequences (ABQ66971-ABQ67178) or their complements (I), also
XX related oligomers, are used to evaluate the methylation status and/or
XX single-nucleotide polymorphisms, in angiogenesis-related genes, for
XX diagnosis and treatment of eye diseases, proliferative retinopathy,
XX neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
XX diabetic retinopathy, macular degeneration caused by neovascularisation,
XX psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
XX Crohn's disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 83391 BP; 24547 A; 665 C; 16953 G; 41209 T; 17 other;
XX
XX Query Match 3.5%; Score 38.4; DB 24; Length 83391;
XX Best Local Similarity 52.5%; Pred. No. 14;
XX Matches 84; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
XX
XX QY 344 TTGCTTTGTAGGGTTTGTATGCGTTGTTTTTTTACTAGGAAGAGCTTGTCTTTTCTATG 403
XX DB 80412 TTTATTTAGAGCGGTGTATAAAGAGGTTTTTTTGTGGAAGGGAATCTTTTTTTATTT 80471
XX
XX QY 404 ACTTTAGTGAAGAGTCTTCCCATGTTTATGAAAAATTCAGAAAATGAGGGATGTAGAAGC 463
XX DB 80472 GGATTTTGTGTAGGTTTTTTTATTATTTTAAAGAGTTGATAGAGAGGGAGGAAAAGG 80531
XX
XX QY 464 CRAAGCACGTGCCCTTAAAGGTCATGATGATGATCAGGAG 503
XX DB 80532 AGAGGAGAGGGATCGGAGGTTGGGGGAGGGGGGGG 80571
XX
XX RESULT 23
XX AAS46333
XX ID AAS46333 standard; DNA; 7667 BP.
XX
XX AAS46333;
XX
XX 18-DEC-2001 (first entry)
XX
```

XX Tumour suppressor gene derived chemically modified sequence #55.
DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
OS Homo sapiens.
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer
XX
PS Claim 1; SEQ ID No 55; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7667 BP; 2446 A; 63 C; 1493 G; 3665 T; 0 other;

Query Match 3.5%; Score 38.2; DB 22; Length 7667;
Best Local Similarity 49.7%; Pred. No. 6.6;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 566 TGCACAGACAGACATTTGATGGGAGCTACAAATCCTGAAACGGGATTAAGGAAAAA 625
DB 2674 TGTATTATAAAAAATTTTATTTAGATTTTAAATTTTGTGGAGAAAAATAGATAAA 2733
QY 626 TGTCTTTCTTAATTCGAAGAAATTAATGACTCTTGATGGTCTAAGAAATCAGAAAGGAGA 685
DB 2734 AATATGATAGTTAATGATTTTAAATAATAATAATTTTAAAGATAGAGATATGAGG 2793
QY 686 CCGAAGCTTCTAAGTAACTAAGCAAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAC 745

DB 2794 TAGAAGTAAATTTATTTATTAAGGAAATGTAGAGAAAATATTGTTATTAGATAAAGGAG 2853
QY 746 CGTGAGGAAAGGAGCT 760
DB 2854 AGAGAGAGAGAGATT 2868
RESULT 24
ABL32294/G
ID ABL32294 standard; DNA; 9415 BP.
XX
AC ABL32294;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 267.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 267; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 9415 BP; 2021 A; 297 C; 2877 G; 4220 T; 0 other;

Query Match 3.5%; Score 38.2; DB 24; Length 9415;
Best Local Similarity 48.8%; Pred. No. 7.1;
Matches 103; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 595 TACAATCCTCGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAGAAATTTAATGAC 654
DB 4049 TAAACCTCCCCACACCTATTACACGAATAAATTTTCTTAATCTACCAAAATAAAAT 3990
QY 655 TGCTTGATGGTCTAAGAAATACAGAAAGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAA 714
DB 3989 TCCCTTCCCACTTCCCTACCAAAATTTATACCAACAATACCAAAACATCCGTAAANAC 3930

CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 113515 BP; 31803 A; 1174 C; 24020 G; 56518 T; 0 other;

Query Match 3.5%; Score 38; DB 24; Length 113515;
 Best Local Similarity 60.8%; Pred. No. 21;
 Matches 62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 343 ATTGCTTTGTAGGGTTGTATGCTGTTGTTTTTACTAGGAGAGCTTGATCTTTCTAT 402
 Db 31032 ATTATTTTATTTTTCGCTGCTGTTGTTTGTATGATGATGTTGTTTTTATTT 31091

QY 403 GACTTTTACTGAAGAGCTCTCCCATGTTTATGAAAAATTCAGA 444
 Db 31092 GAAGTTTTTGAAGAGTTATTTTTTTTAAAGGAAAAATTAATA 31133

RESULT 29
 ABL33122/C
 ID ABL33122 standard; DNA; 7690 BP.

XX ABL33122;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1095.

XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; anti-anaemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX W0200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPT; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 1095; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 7690 BP; 1804 A; 225 C; 2151 G; 3510 T; 0 other;

Query Match 3.5%; Score 37.8; DB 24; Length 7690;
 Best Local Similarity 48.0%; Pred. No. 8.5;
 Matches 108; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 672 AATCAGAAAGGAGACCGGAAGCTTCTAGTAATCAAGAAAAAATCTTGACAAATAGCTTT 731
 Db 2689 AATCAATTTAAACTAAATAATATATAAAACCTTTATTTCTTAATTTTAACTT 2630

QY 732 TCCATAGAGAAACCTTGAGGAAGAGTCCGATACCAATAATCCGATTAAAAAAGGAG 791
 Db 2629 TCTAAAAAATAAATAATATATAAATAAACAATAATACGTTCTATAACAATTTTAA 2570

QY 792 TGCTTACACAGCCATATCAGTGGAGCAACCTTCCCTTTGATTTTGATATGACAGAA 851
 Db 2569 AACTTAAAAAACCAATTTAATAATAAACTTCTATCTCCCTTTTAAACAATACAA 2510

QY 852 GATGACAAAGAGTAAAGTTTTTGAAAAATTTCTAAAAATAGAAATGC 896

Db 2509 AACAAAAACGCAAAAAACAATAATAACTAATACTTCGAAATCC 2465

RESULT 30
 AAC40730/C
 ID AAC40730 standard; DNA; 870 BP.

XX AAC40730;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 29351.

XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132048.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138844.
PR 14-JUN-1999; 99US-0139111.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 17-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140822.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151203.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 3.4%; Score 37.4; DB 21; Length 870;		
Best Local Similarity 51.5%; Pred.No. 4.8;		
Matches	86; Conservative	0; Mismatches 81; Indels 0; Gaps 0;
OY	30 ATTCTCTTGATGTAAAGGGGAATTGATTCTAACAGACCTTGGTCGAGCATCCATTCA	89
Ddb		
168 ATTTCTCCGGTGGTGAGCGGCCTACCAGAAGTACTAGAGGAGCGAGTCATTTTG	109	
OY	90 ATTTGAAGTGAAGTCTTACCAAGTTCATCAAATGTCTATTGAAAAATAAAGTCTCTTGTG	149
Ddb		
108 TTTTGGTTTTGAGTCTCTGTTTTTAGGGGAAGAGCAAAATGTGAAGACAGAGAAGAGAG	49	
OY	150 GAGTGTGATGTATTGAAGGGAAGACAATAATCTGATCCAATTCTAAA	196
Ddb		
48 ACTTTGTTGTTTGTAGAGGAGAAGAATCTGATCCAAGTTATAA	2	
RESULT 31		
ABL3J359		
ID	ABL3J2359 standard; DNA; 6237 BP.	
XX	AC ABL3J2359;	
XX	XX	
DT	26-MAR-2002 (first entry)	
DE	Human immune system associated gene SEQ ID NO: 332.	
XX	XX	
XX	Human; immune system disease; cytosine methylation; antiasthmatic;	
KW	antiarteriosclerotic; antianaemic; cytostatic; neurotrophic;	
KW	neuroprotective; anti-HIV; anticoagulant; ophthalmological;	
KW	antirheumatic; antiarthritis; antididiabetic; antipsoriatic;	
KW	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;	
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;	
KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;	
gene; ds.		
XX	XX	
OS	Homo sapiens.	
XX	WO200200928-A2.	
PX	03-JAN-2002.	
PD	02-JUL-2001; 2001WO-EF07537.	
PF	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
PA	(EPIG-) EPIGENOMICS AG.	
PI	Olek A, Piepenbrock C, Berlin K;	
XX	WPI; 2002-130909/17.	
XX	Nucleic acid comprising fragment of chemically modified gene, useful	
PT	for diagnosis and treatment of diseases associated with abnormal	
PT	cytosine methylation -	
XX	Claim 1; SEQ ID NO 332; 32pp + Sequence Listing; German.	
PS	XX	
XX	The present invention provides a number of human immune system associated	
CC	genes which are modified by the methylation of cytosines. The sequences	
CC	can be used in the diagnosis and treatment of immune system disorders,	
CC	including eye diseases such as retinopathy, neovascular glaucoma and	
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid	
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,	
CC	rheumatoid arthritis, psoriasis and inflammatory ulcerative bowel	
CC	diseases. The present sequence is a gene of the invention.	

XX	Seq	Sequence 6237 BP; 2071 A; 77 C; 1289 G; 2800 T; 0 other;
	Query Match	3.4%; Score 37.4; DB 24; Length 6237;
	Best Local Similarity	50.3%; Pred. No. 10;
	Matches	92; Conservative 0; Mismatches 91; Indels 0; Gaps
OY	342	GATTGCTTTGACGGTTCTGATCGCTTTCTTTTACTAGAGAGCTGTACTTTTCTA 401
Db	1882	GAAGGATATATTTTTTTTCATAGGAAGTTTTTTTAAAGGATATATTTTTTTT 1941
OY	402	TGACTTTAGTGAAGAGCTTCCCATGTTTATGAAAAATTCAGAAATCAGGAGTGTAGAA 461
Db	1942	ATTAATTAATTTATTTATTAAGAAATATTAATAAAGTTTTAGATAAGTTAAGTTGAT 2001
OY	462	GCCAGGACGTCGCCCTAAAGGTCATGATGACATCAGGATCAGGAGTTAAGGAGAAATTAICTT 521
Db	2002	TTTAAATTTGATGATGAAAAAATTAATGTAAAGATAGTTAATAATTTAAAGTTATAGT 2061
OY	522	TAT 524
Db	2062	AAV 2064
RESULT 32		
ABL32530		
ID	ABL32530	standard; DNA; 7644 BP.
XX	AC	ABL32530;
XX	AC	ABL32530;
XX	DT	26-MAR-2002 (first entry)
XX	XX	Human immune system associated gene SEQ ID NO: 503.
DE	XX	Human; immune system disease; cytosine methylation; antiasthmatic;
XX	XX	antiarteriosclerotic; antiandemic; cytostatic; neurotropic;
KW	KW	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW	KW	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW	KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	KW	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW	KW	gene; ds.
XX	XX	Homo sapiens.
XX	OS	
XX	XX	WO200200928-A2.
XX	PN	
XX	XX	03-JAN-2002.
XX	PD	
XX	PF	02-JUL-2001; 2001WO-EP07537.
XX	PF	
XX	PR	30-JUN-2000; 2000DE-1032529.
XX	PR	01-SEP-2000; 2000DE-1043826.
XX	PR	
XX	PA	(EPIG-) EPIGENOMICS AG.
XX	XX	
XX	PI	Olek A, Piepenbrock C, Berlin K;
XX	XX	
DR	WI	2002-130909/17.
XX	XX	
XX	PT	Nucleic acid comprising fragment of chemically modified gene, useful
XX	PT	for diagnosis and treatment of diseases associated with abnormal
XX	PT	cytosine methylation
XX	XX	
XX	PS	Claim 1; SEQ ID NO 503; 32pp + Sequence Listing; German.
XX	XX	
CC	CC	The present invention provides a number of human immune system associated
CC	CC	genes which are modified by the methylation of cytosines. The sequences
CC	CC	can be used in the diagnosis and treatment of immune system disorders
CC	CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

```
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7644 BP; 2187 A; 53 C; 1625 G; 3779 T; 0 other;

Query Match      3.4%; Score 37.4; DB 24; Length 7644;
Best Local Similarity 60.2%; Pred. No. 11;
Matches 62; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 338 ATCAGATTGCTTGAGGCTTGATGCTGTTGTTTACTAGAGAGCTGTACTATTT 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2970 ATTAGTTGCTTTATGCTTTATGCTTTGTTGTTGTTAGAGTATTCCTTTGATTTT 3029
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 398 TCATGACCTTACTGAGAGCTCCAGATGTTTATGAATAATT 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3030 TGTATTATTAGATGAAGTTTGTTGTTGTTGATTAATGAATT 3072
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 33
AAC36518/c
ID AAC36518 standard; DNA; 522 BP.
XX
AC AAC36518;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14105.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
FD 06-SEP-2000.
XX
FF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130519.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
```

```
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149302.
PR 23-AUG-1999; 99US-0149330.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151530.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159384.
PR 13-OCT-1999; 99US-0159385.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.4%; Score 37.2; DB 21; Length 522;
Best Local Similarity 51.9%; Pred No 4.5; 78; Indels 0; Gaps 0;
Matches 84; Conservative 0; Mismatches 0;

QY 30 ATTTCTCTTGATTGTAAGGGAATTGATTCACAGACGCTGGTCGAGCATCCATTCTA 89
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 168 ATTTCTTCCGGTGGGAGGCGCCCGTACCAAGGATCTAGAGGAGGAGTCAATTTG 109
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 90 ATTTGAAGTGAGTTCACGAGTTCATCAAAATGCTTTATGAAAATAAAGTCTCTTGTG 149
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 108 TTTTGGTTTGGATCTCTGTTTATAGGGAAGAGGAATGTGAAGACAGAGGAAGAGAG 49
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 150 GAGTTGATGTATTGAAGGGAAGCAATAATCTCATCCAATT 191
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 ACTTTGTTGTTGTTTATAGGAGAGAGGAATCTGTATCCAAGT 7
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 34
AAC51121/c
ID AAC51121 standard; DNA; 869 BP.
AC AAC51121;
XX
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67351.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
```

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142058.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149529.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to
CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3345 BP; 1108 A; 485 C; 551 G; 1201 T; 0 other;

Query Match 3.4%; Score 37.2; DB 22; Length 3345;
Best Local Similarity 47.1%; Pred. No. 9;
Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 574 AGACAAGTATTGATGGGGAGCTACAAATCCTGAACGGGATAATAGAAAATGTGTTTC 633
DB 2911 ATAAAAAATTCATATAGATCTCGTTCCCTCATCTGTATACAAATAATTAAGTTAT 2970
QY 634 TAATTCGAAGAATTATGACCTGCTTGATGGTCTAAGAAATCAGAAGGAGACCGAAAGC 693
DB 2971 GAACACCCCAATATTAAACAAACACCTTTTACTCTAACTCTAAATAAAACCACTTT 3030
QY 694 TTCTAAGTAAATCAAGAAAAATCTTGGCAATAGCTTTTCCATAGAGAAAAACGGTGAG 753
DB 3031 GTTTAATTAAGTGAATTTCTTCTCATCATATTTTCTGATAAAAAATAACCTGGGAC 3090
QY 754 AAGGAGTCGATACCAATAATCCGATTAAAAAGGGAGTCTTACACACGCCATATCAGT 813
DB 3091 AATCAATTTTCCCGAGTATTATTAATAATAACATCATGACAAAAATTAATTTGAGT 3150
QY 814 GG 815
DB 3151 TG 3152

RESULT 37
ABL32802
ID ABL32802 standard; DNA; 6113 BP.
XX
AC ABL32802;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 775.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosine methylation; antiasthmatic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 775; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6113 BP; 1487 A; 157 C; 1487 G; 2982 T; 0 other;

Query Match 3.4%; Score 37.2; DB 24; Length 6113;
Best Local Similarity 53.4%; Pred. No. 11;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 860 AAGAGTAAGTTTTTGAAGAAATCTAAATAGAAATGCAATTTGTCGAGCTAAAGCTTGC 919
DB 337 AAAAGAAGATTCGTAAGAAAGGTTTTAAAGTGTATGTTAGGTAGAGAAAAAATTTG 396
QY 920 TTCTCTTTATTTCTTTCTTCTAGCAATGATTCGGTAGTAGGATATGGGTCGAGTATGCA 979
DB 397 TTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 456
QY 980 CGTTGGAGTATTCGGTGTCTCGAA 1005
DB 457 TTTATACGTGTTTCGGTGTGTGTGAA 482

RESULT 38
ABL33720
ID ABL33720 standard; DNA; 7190 BP.
XX
AC ABL33720;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1693.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytosine methylation; antiasthmatic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
XX for diagnosis and treatment of diseases associated with abnormal
XX cytosine methylation

PS Claim 1; SEQ ID NO 1693; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 7190 BP; 1434 A; 310 C; 2114 G; 3261 T; 71 other;
 Query Match 3.4%; Score 37.2; DB 24; Length 7190;
 Best Local Similarity 49.5%; Pred. No. 12;
 Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 QY 837 TTGATAATGACAGAGATGAACAAAGAGTAAGTTTTCGAAAAATCTTAAATAGAAATGC 896
 DB 6770 TAGATGANTAGTTATGAAGATAATATTGNATGTTTINAGAAGAATGTNAGATTATTTTT 6829
 QY 897 ATTGTGTCGAGCTAAAGCTTGCTCTCTTTATTTTCCTTTTGTAGAAATGATTCGGTAGT 956
 DB 6830 GGAAGTGTGGGGAAGGTGGAAGTGAATTTATGATTGTTAGTAGAAAGTTACGTTAGA 6889
 QY 957 AGGAATATGGGTGCGATGATGACGTTGGAGTGATTCGCTGCTCTGCAAGAACAGGGA 1016
 DB 6890 GTTTATATAGTTTGTATTGATAGAGTGGGAGGGGTGTAGGGGAAGAGAGGTAAA 6949
 QY 1017 GT 1018
 DB 6950 TT 6951
 RESULT 39
 ABV52129/c
 ID ABV52129 standard; cdna; 619 BP.
 XX
 AC ABV52129;
 XX
 DT 17-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cdna 52120.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US051171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207434P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 PT

PS Claim 1; Page 10116; 11750pp; English.
 XX The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 619 BP; 265 A; 95 C; 107 G; 149 T; 3 other;
 Query Match 3.4%; Score 37; DB 23; Length 619;
 Best Local Similarity 60.4%; Pred. No. 5.4;
 Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
 QY 330 CAATAGGCGATCAGATTGCTTTGTAGGGTTTCTATCGGTGTTTCTTACTAGGAAGAGCTT 389
 DB 427 CATAACGCTTAGAAGCTTTCTTAAGTTTCTTCTGTGTGTTTCTTCTCCCATAGATT 368
 QY 390 GTACTTTTCTTCTATGACTTTAGTAGAGAGTCTTCCCATGTTT 430
 DB 367 TTTGTTTTCATATTTGTTAGTCAAGCTTTGTCATTATGTTT 327
 RESULT 40
 ABV31005
 ID ABV31005 standard; DNA; 1134 BP.
 XX
 AC ABV31005;
 XX
 DT 05-JAN-2000 (first entry)
 XX
 DE Partial dnaN gene.
 XX
 KW Gram positive bacteria; dnaE, dnaX, dnaB; dnaC; dnaN; dnaG; helicase;
 KW alpha subunit; DNA polymerase III holoenzyme; gamma subunit; tau subunit;
 KW clamp loader; glue protein; replication; antibiotic; ss.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1134
 FT /*tag= a
 FT /note= "Dna N gene product"
 XX
 PN W09937661-A1.
 XX
 PD 29-JUL-1999.
 XX
 PF 25-JAN-1999; 99WO-US01547.
 XX
 PR 27-JAN-1998; 98US-0074522.
 PR 22-JUL-1998; 98US-0093727.
 XX
 PA (UYRQ) UNIV ROCKEFELLER;
 XX
 PI O'Donnell ME, Zhang D, Whipple R;
 XX
 DR WPI; 1999-590685/50.
 DR P-PSDB; AAY49071.
 XX
 XX New isolated dnaE, dnaX and dnaB genes from Gram positive bacteria,
 PT used to develop screening assays for identifying antibiotic compounds -
 PT

XX Example 8; Page 30; 132pp; English.

XX This sequence is the partial dnaN gene of *Staphylococcus aureus*. The

XX invention relates to a number of isolated DNA molecules from Gram

XX positive bacterium, corresponding to dnaE (AAZ31001), dnaX

XX (AAZ31002), and dnaB (AAZ31003). The polC, dnaE and dnaB genes

XX (AAZ31004-231006) are also identified. The dnaE gene corresponds to the

XX alpha subunit of the *Escherichia coli*, DNA polymerase III holoenzyme,

XX dnaX corresponds to the gamma and tau subunits, and dnaB corresponds to

XX the helicase. The alpha subunit is the actual DNA polymerase, the gamma

XX complex forms the clamp loader and tau is a "glue protein". DnaX encodes

XX both gamma and tau, tau is the product of the full gene, while gamma is

XX the product of the first two thirds of the gene. DnaN forms the beta

XX subunit which forms the sliding clamp, and dnaG encodes a primase. The

XX DNA sequences of the invention can be used to identify agents that

XX inhibit or promote DNA replication by acting on various parts of the gram

XX positive bacterial DNA polymerase holoenzyme. The products and methods of

XX the invention can be used for identifying pharmacological agents or lead

XX compounds for agents active at the level of a replication protein

XX function, particularly DNA replication. The agents identified can be used

XX as antibiotics.

XX

SQ Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 other;

Query Match 3.4%; Score 37; DB 20; Length 1134;

Best Local Similarity 50.3%; Pred. No. 6.8;

Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 34 CTCCTGATTTCTAAGGGGAATTCATCTTAACAGACCTGGTCGAGCATCCATCTTAATTT 93

DB 834 CTCCTGATTTCTAAGGGGAATTCATCTTAACAGACCTGGTCGAGCATCCATCTTAATTT 93

QY 94 TGAAGTGAGTCTTACAGGTCATCAAAATGCTTATTGAAAATAAAGTCTCTTGTGGAGT 153

DB 894 TGTGGAATTAATCTCTTACATCAACAGAAATTTGGTACTGTAAGAAGAAGAGTGTGATGCAAA 953

QY 154 TGAGTATTGAAGGGAAGACGAATATCTGATCCAAATCTTAAAGGGAGACACGATCGTT 213

DB 954 CGATGTTGAAGGTGGTGCTGAAATTTTCATTCAACTCTAAATATATATGATGATGCTTT 1013

QY 214 A 214

DB 1014 A 1014

RESULT 41

AAF54735

XX AAF54735 standard; DNA; 1134 BP.

XX AAF54735;

XX 15-MAY-2001 (first entry)

XX Nucleotide sequence of a dnaN polypeptide.

XX dnaE; Gram positive bacteria; polC; dnaE; hola; holaB; dnaX; dnaN; ssb;

XX dnaG; dnaB; antibiotic; replication; cell growth; cell death;

XX bacterial infection; ss.

XX *Staphylococcus aureus*.

XX Key Location/Qualifiers

XX CDS 1..1134

XX /tag= a

XX /product= "dnaN protein"

XX W0200109164-A2.

XX 08-FEB-2001.

XX 28-JUL-2000; 2000WO-US20666.

XX

PR 29-JUL-1999; 99US-0146178.

XX (UYRQ) UNIV ROCKEFELLER.

XX O'Donnell ME, Bruck I, Zhang D, Whipple R;

XX WPI; 2001-147453/15.

XX P-PSDB; AAB31935.

XX Isolated DNA molecule from a Gram positive bacterium encoding DNA

XX replication proteins used to identify compounds which have antibiotic

XX activity -

XX Disclosure; Page 39-40; 239pp; English.

XX The present sequence encodes a dnaN polypeptide. The specification

XX describes DNA molecules from Gram positive bacteria, which comprise

XX a coding region from a polC, dnaE, hola, holaB, dnaX, ssb, dnaG

XX or a dnaB gene. These sequences encode proteins that replicate the

XX chromosome of Gram positive bacteria. They are used for sequencing the

XX CC and amplification of DNA and in drug discovery to identify compounds

XX which have antibiotic activity through interference with replication.

XX They are used in methods for identifying compounds that are active at

XX the level of DNA replication and result in arrest of cell growth or

XX cell death of bacteria to treat bacterial infections in animals.

XX

SQ Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 other;

Query Match 3.4%; Score 37; DB 22; Length 1134;

Best Local Similarity 50.3%; Pred. No. 6.8;

Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 34 CTCCTGATTTCTAAGGGGAATTCATCTTAACAGACCTGGTCGAGCATCCATCTTAATTT 93

DB 834 CTCCTGATTTCTAAGGGGAATTCATCTTAACAGACCTGGTCGAGCATCCATCTTAATTT 93

QY 94 TGAAGTGAGTCTTACAGGTCATCAAAATGCTTATTGAAAATAAAGTCTCTTGTGGAGT 153

DB 894 TGTGGAATTAATCTCTTACATCAACAGAAATTTGGTACTGTAAGAAGAAGAGTGTGATGCAAA 953

QY 154 TGATGATTTGAAGGGAAGACGAATATCTGATCCAAATCTTAAAGGGAGACACGATCGTT 213

DB 954 CGATGTTGAAGGTGGTGCTGAAATTTTCATTCAACTCTAAATATATATGATGATGCTTT 1013

QY 214 A 214

DB 1014 A 1014

RESULT 42

AAH32611/C

XX AAH32611 standard; cDNA; 1075 BP.

XX AC AAH32611;

XX 10-AUG-2001 (first entry)

XX Human secreted protein gene 27 cDNA clone HJAAJ58, SEQ ID NO:100.

XX Human; secreted protein; proliferative disorder; cancer;

XX fetal abnormality; developmental abnormality; haematopoietic disorder;

XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

XX inflammation; allergy; neurological disorder; Alzheimer's disease;

XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;

XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

XX cardiovascular disorder; angioedema; kidney disorder;

XX gastrointestinal disorder; pregnancy-related disorder; tumour;

XX endocrine disorder; infection; wound healing; vulnerability;

XX cell culture; chemotaxis; food additive;

XX binding partner identification; ss.

XX Homo sapiens.

XX

CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.

XX Sequence 1724 BP; 433 A; 242 C; 279 G; 770 T; 0 other;

Query Match 3.4%; Score 36.8; DB 22; Length 1724;
Best Local Similarity 52.6%; Pred. No. 9;
Matches 103; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

Qy 694 TTCTAAGTAATCAAGAAAAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGA 753

Db 1069 TTGTTTGTAAAAAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1010

Qy 754 AAGGAGTCGATACCAATAATCCGATTTAAAAAGGAGTGCTTTACACACGCCATATCACT 813

Db 1009 AAGACTCAATTAATGAATCAGAAATGAAAGGAGACATGCAATTCATACACACAGA 950

Qy 814 GGAGCAACCTCCCTTTGATTTCTTTGATAATGACAGAGA-TGAACAAGAGTAAGTTT 872

Db 949 AATCAAGAATCATTTGGAGTCTTATACATCAACCATGAGTCAACAATTTGAAAGCCT 890

Qy 873 TGAATAATTCATAAT 888

Db 889 AGTAAAAATGTATAAT 874

RESULT 44

AAK68427

ID AAK68427 standard; DNA; 1761 BP.

AC AAK68427;

DT 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23239.
XX Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX W0200157182-A2.

PN 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216680.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234277.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0234584.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246526.

PR 22-AUG-2000; 2000US-0225868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 29-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX

PS Disclosure: SEQ ID NO 23240; 307lpp + Sequence Listing: English.

CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.

XX Sequence 1761 BP; 809 A; 297 C; 253 G; 402 T; 0 other;

Query Match 3.4%; Score 36.8; DB 22; Length 1761;

Best Local Similarity 52.6%; Pred. No. 9;

Matches 103; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

QY 694 TTCTAAGTAATCAAGAAATCTTCGACAAATAGCTTTCCATAGAGAAACCGTGAGGA 753

Db 699 TTGTTTGTAAAAAATAACAAATTAACAAATGTTTACCTAAACAAAGTAAGAAAAAG 758

QY 754 AAGGAGTCGATACCATTAATCCGATTAAAAAGGGAGTGCTTACACAGCCATATCAGT 813

```

Db 759 AAAGACTCAATAAATGAATCAGACATCAAAAAAGGAGACATGACAAATGATACACAGA 818
QY 814 GGAGCAACCTTCCCTTTGATTCCTTTGATAATGACAGAGA-TGAACAAAGAGTAAGTTTT 872
Db 819 AATACAAGAAGATCATGGAGCTCTATTACAATCAACCATGAGTCAACAAATTTGGAAGCCT 878
QY 873 TGAATAAATTTCTAAAT 888
Db 879 AGTAATAATGTATAAT 894

RESULT 46
ABQ67075
ID ABQ67075 standard; DNA: 7857 BP.
XX
AC ABQ67075;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 105.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EPI4320.
XX
PR 06-DEC-2000; 2000DE-1061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated
PT angiogenesis-associated genes, useful for determining methylation
PT status, e.g. in diagnosis or treatment of cancer
XX
PS Claim 1; SEQ ID NO 105; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7857 BP; 2110 A; 218 C; 2100 G; 3429 T; 0 other;

Query Match 3.4%; Score 36.8; DB 24; Length 7857;
Best Local Similarity 54.4%; Pred. No. 16;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 829 TTGATCTTTGATPATGACAGAGATGACAAAGAGTAAGTTTTTGAAATAATCTAAAT 888
Db 3507 TTAGATGATTTTAAAGACGGGAGATATTTTATTTTAGTATTTTATTTTAAAAA 3566

```

```

QY 889 AGAATGTCATTTGTCGAGCTAAAGCTTCTCTCTTATTTTCTTTTGTAGATGAT 948
Db 3567 AAAAAAATTTTCTTTAGCTGTAGTGGTTATGTTTCTATTTTATTTTGAAGAT 3626
QY 949 TCGGTAGTAGGATAT 964
Db 3627 AAGTAGTGGATTAT 3642

RESULT 47
ABA03041/C
ID ABA03041 standard; DNA: 2944528 BP.
XX
AC ABA03041;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes EGD-e genome sequence.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.
XX
KW Listeria monocytogenes.
XX
OS WO200177335-A2.
XX
PN 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP ) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides
XX
PS Claim 1; SEQ ID No 1; 192pp; French.
XX
CC The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (AB47297-AB50149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccines compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match 3.4%; Score 36.8; DB 24; Length 2944528;

```

CC	in gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the protein encoded by
CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB35893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
CC	of the present invention.
XX	
SQ	Sequence 771 BP; 246 A; 131 C; 124 G; 264 T; 6 other;
	Query Match 3.4%; Score 36.6; DB 22; Length 771;
	Best Local Similarity 65.1%; Pred. No. 7.5;
	Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0
QY	816 AGCAACCCCTCCCTTGATTCCTTTGATAATGCAGACAAGAATGAACAAAGAGTAAAGTTTGA 875
Db	567 AGCCACAAAAGCCTTTGAAGGCTTTGATTATTCAGAGAAGATGACCAAAGAGTCTCTTTTATTT 508
QY	876 AAAATCTCAAAATAGAATGCAT 898
Db	507 AGGCTACTCGAATATATTTTTAT 485

RESULT 49	
AAV22649/c	
ID	AAV22649 standard; cDNA; 2098 BP.
XX	
AAV22649;	
XX	
17-AUG-1998	(first entry)
XX	
Homo sapiens soluble kuzbanian (kuz) gene.	
XX	
kuzbanian; kuz; neurogenic; KUZ protein; neuronal partitioning;	
development; NOTCH protein processing; regulation; cell function;	
signal transduction pathways; screening; receptor binding;	
metalloprotease; soluble; ds.	
XX	
Homo sapiens.	
OS	
Key	Location/Qualifiers
FD	2..2098
CDS	/*tag= a
FT	/product= KUZ protein
FT	
FT	
PN	WO9808933-A1.
XX	
05-MAR-1998.	
XX	
27-AUG-1997;	97WO-US15099.
XX	
23-JUL-1997;	97US-0019390.
PR	
29-AUG-1996;	96US-0019390.
PR	
XX	
(REGC) UNIV CALIFORNIA.	
PA	{OYIA } UNIV YALE.
XX	
Pan D, Rooke J, Dubin GM, Xu T, Yavari R;	
PI	
XX	
WPI; 1998-179428/16.	
DR	
P-PSDB; AAW56133.	
DR	
XX	
New KUZ polypeptides, members of the ADAM family of metalloprotease	
- useful in neural partitioning and development	
XX	
Claim 11; Pages 37-38; 58pp; English.	
XX	
The sequence is that of the human soluble (lacking a	
transmembrane domain) kuzbanian gene, a neurogenic gene.	
CC	
The KUZ protein that it encodes is involved in neuronal	
CC	
partitioning and development. It is also involved in processing	
CC	

CC of the NOTCH protein by cleaving the C-terminal portion. The
CC KUZ protein can be used to screen for compounds that alter
CC binding of KUZ to its receptor or the cleavage of the NOTCH
CC protein, hence acting to regulate NOTCH signal transduction
CC pathways and regulate cell functions.

XX
XX Sequence 2098 BP; 655 A; 353 C; 380 G; 710 T; 0 other;
SQ
Query Match 3.4%; Score 36.6; DB 19; Length 2098;
Best Local Similarity 65.1%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 816 AGCACCTTCCCTTTGATTTGATGATGACAGAGATGACAGAGATGATTTTGA 875
Db 1136 AGCCACAAAGCTTTGAGGTTTGATTTTCAGAGAGATGACAGAGATTTTATTT 1077
QY 876 AAAATTTCTAAATAGAAATGCAT 898
Db 1076 AGGCTACTGAAATATTTATTTAT 1054

RESULT 50
AAH17997/c
ID AAH17997 standard; cDNA; 2333 BP.

XX AC AAH17997;
XX AC
XX 26-JUN-2001 (first entry)
XX Human cDNA sequence SEQ ID NO:17797.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 17797; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 2333 BP; 699 A; 427 C; 416 G; 791 T; 0 other;
SQ
Query Match 3.4%; Score 36.6; DB 22; Length 2333;
Best Local Similarity 65.1%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 816 AGCAACCTTCCCTTTGATTTGATGATGACAGAGATGACAGAGATGATTTTGA 875
Db 567 AGCCACAAAGCTTTGAGGTTTGATTTTCAGAGAGATGACCAAGAGTTCTTTTATTT 508
QY 876 AAAATTTCTAAATAGAAATGCAT 898
Db 507 AGGCTACTGAAATATTTATTTAT 485

Search completed: January 9, 2003, 10:51:58
Job time : 2027 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 05:36:27 ; Search time 2610 Seconds
(without alignments)
12109.443 Million cell updates/sec

Title: US-09-438-185a-1_COPY_1199590_1200675

Perfect score: 1086
Sequence: 1 ttggcaagtagctcaaac.....gtctctgcagtagtgac 1086

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

GenEmbl.:

- 1: gb_ba.:
- 2: gb_htg.:
- 3: gb_in.:
- 4: gb_ov.:
- 5: gb_ov.:
- 6: gb_ov.:
- 7: gb_ph.:
- 8: gb_pl.:
- 9: gb_pr.:
- 10: gb_ro.:
- 11: gb_sy.:
- 12: gb_sy.:
- 13: gb_un.:
- 14: gb_vl.:
- 15: em_ba.:
- 16: em_fun.:
- 17: em_hum.:
- 18: em_in.:
- 19: em_mu.:
- 20: em_om.:
- 21: em_or.:
- 22: em_ov.:
- 23: em_pat.:
- 24: em_ph.:
- 25: em_pl.:
- 26: em_ro.:
- 27: em_sts.:
- 28: em_un.:
- 29: em_vl.:
- 30: em_htg_hum.:
- 31: em_htg_inv.:
- 32: em_htg_other.:
- 33: em_htg_mus.:
- 34: em_htg_pln.:
- 35: em_htg_rod.:
- 36: em_htg_mam.:
- 37: em_htg_vrt.:
- 38: em_sy.:
- 39: em_htgo_hum.:
- 40: em_htgo_mus.:
- 41: em_htgo_other.:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	1086	100.0	1089	6	AX349679	Sequence
C 2	1086	100.0	10847	1	AE001685	Chlamydia
C 3	1086	100.0	22589	1	AE002240	Chlamydia
C 4	1086	100.0	325855	1	AF002548	Chlamydia
C 5	72.4	6.7	77218	6	166494	Sequence 14
C 6	46.8	4.3	151120	2	AC015814	Homo sapi
C 7	46.8	4.3	193804	9	AC084882	Homo sapi
C 8	46.6	4.3	1642	3	AF135186	Caenorhab
C 9	45.6	4.2	1531	3	AF119388	Caenorhab
C 10	45.6	4.2	41246	3	CEK08F8	Caenorhab
C 11	45.6	4.2	109206	2	AC121238	Medicago
C 12	45	4.1	1141	6	AX083744	Sequence
C 13	45	4.1	172246	9	AC024941	Homo sapi
C 14	45	4.1	180563	2	AC034170	Homo sapi
C 15	44.8	4.1	12029	3	AE001431	Plasmodiu
C 16	44.2	4.1	783	9	HSA343706	Homo sapi
C 17	44.2	4.1	234112	3	PFMAL4P2	Plasmodiu
C 18	44	4.1	167257	9	AC008008	Homo sapi
C 19	44	4.1	198230	9	AC092562	Homo sapi
C 20	43.8	4.0	1141	6	AX083744	Sequence
C 21	43.4	4.0	6089	6	AX344846	Sequence
C 22	43.2	4.0	88866	2	AP001943	Homo sapi
C 23	43.2	4.0	146671	9	AC093810	Homo sapi
C 24	42.8	3.9	158066	9	AC107307	Homo sapi
C 25	42.8	3.9	159716	2	AC025807	Homo sapi
C 26	42.8	3.9	186158	9	CNS018F0	Human chr
C 27	42.8	3.9	188385	2	AC112384	Rattus no
C 28	42.4	3.9	159035	9	AC092564	Homo sapi
C 29	42.2	3.9	2706	1	AF268060	Candidatu
C 30	42.2	3.9	126010	9	AC010342	Homo sapi
C 31	42.2	3.9	141635	2	AL080007	Homo sapi
C 32	42.2	3.9	132168	2	AL807375	Mus muscu
C 33	42.2	3.9	202801	2	AL831718	Mus muscu
C 34	42.2	3.9	266050	1	AP000988	Apicomplex
C 35	42	3.9	98637	9	AC004854	Homo sapi
C 36	42	3.9	124767	2	AC108637	Rattus no
C 37	41.8	3.8	363	1	MC445	M. capricolu
C 38	41.8	3.8	76141	2	AC024393	Homo sapi
C 39	41.8	3.8	115666	2	AC105744	Oryza sat
C 40	41.8	3.8	126371	9	AL391883	Human DNA
C 41	41.8	3.8	156060	2	AC004153	Plasmodiu
C 42	41.8	3.8	157515	2	AL391598	Homo sapi
C 43	41.8	3.8	194247	2	AL358253	Homo sapi
C 44	41.8	3.8	200279	2	AC109165	Mus muscu
C 45	41.6	3.8	27785	2	AC116978	Dictyoste
C 46	41.6	3.8	69674	9	AL136130	Human DNA
C 47	41.6	3.8	103337	2	AC115405	Rattus no
C 48	41.6	3.8	110000	2	AL353694_1	Continuation (2 of
C 49	41.6	3.8	148549	2	AL354818	Homo sapi
C 50	41.6	3.8	215896	2	AC046180	Homo sapi
C 51	41.4	3.8	161885	2	AC026518	Homo sapi
C 52	41.4	3.8	181200	2	AC099365	Rattus no
C 53	41.4	3.8	345777	2	AC078629	Homo sapi
C 54	41.2	3.8	14514	2	AC115610	Dictyoste
C 55	41.2	3.8	147192	2	AC116925	Dictyoste
C 56	41.2	3.8	153133	2	AC106207	Rattus no
C 57	41.2	3.8	170668	2	AC091787	Rattus no
C 58	41.2	3.8	197127	2	AC098253	Sequence
C 59	41.2	3.8	349980	6	AX344553	Arabidops
C 60	41	3.8	1202	8	AF308778	Sequence
C 61	41	3.8	9504	6	AX323794	Sequence
C 62	41	3.8	40392	3	U21308	Caenorhab
C 63	41	3.8	137830	2	AC024346	Homo sapi
C 64	41	3.8	151225	9	AC012625	Homo sapi
C 65	41	3.8	152679	2	AC016082	Homo sapi


```
/db_xref="GI:4377383"
/translation="MSKIVYKFGTSLATAENICLVCDIICKDKRPFVYVSAIAGVTD
LIVDCSSLREREREKTEGHEEIVKNAIPFPVTSWTSRLLPYLQHLLEISDLDF
ARILSLGDIISASVRAVCTRWGDLGFLAERSVILTDSDVSRASPNLDMKAHWHOL
ELNPSYITIOGIFASNGITGVLLGRGSDYSATLIAELARATEVRYTDVNGIYWD
PKVISAQRIEPLSEFEMONLASFGAKVLYPPMLPCPCMRGPIFVTSIDPEKGGTW
VYAVDSQVSEPRKALSLSQYQFCSDVYTLVGGGLEILGIESHGIDPELMLIAQ
NUNVGFMDDDIISQEAQELVDVLSLSSVTRLHHSVALITMIGNLSSPKVSVTIE
KRVGFQGVFCFQSSMALSFVVASLAEGLIEELNHLVYKQKAIAT"
5162..6037
/gene="dapa"
/note="Cpnl050"
5162..6037
/gene="dapa"
/codon_start=1
/transl_table=1
/product="dihydrodipicolinate Synthase"
/protein_id="AAD19187.1"
/db_xref="GI:4377384"
/translation="MHLLTATVPFPNGTIDFASLERLLSFQDAGVNGVVLGSGE
GLSLTKKQKALICFACDLQKVPFLVGTSGTLLEEVLDWIHFCNDLPISGLMTTPI
YTKPKLQGLIWFFAVLNAAKHPAILYNIPSRATPLVLDTVKALAHHPQFLGKDSG
GSVEFQSKYTIAPHIQLYCGDDVFWSEMAACCAHGLISVLSNAPPEAREYVLNPOE
QDYSIMMETCRWYVTTNPIGIKAILAYKKAITHAQLRLPLSIEDFLENVSPVES
MLAWPKLRTSVFYS"
6198..6509
/gene="Cpnl051"
6198..6509
/gene="Cpnl051"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19188.1"
/db_xref="GI:4377385"
/translation="MTPRSIQOLHLKTIIDPVKIKSPVTTKKSSFROSLLRLFLELFW
MFLYICIRSFCHVHIATFCRGLILFLATLFLSMICILHFTIPLWICKEDPRIIRKN
K"
6653..7408
/gene="Cpnl052"
6653..7408
/gene="Cpnl052"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19189.1"
/db_xref="GI:4377386"
/translation="WKYNSREKIKSALRICSSVCITVPRNPFSLSCYDKIFYLSQVY
FNGPNSIGCRSFCFFRGKTEVETKEVKKIDKIEPSLEGNDPVKVAESPRAALE
SLSSQSSIGNLCAISNLDQMLSRNFSKEIINGSLTFTSKSDCAESEPFRYTAG
VLGRSKLAGSLYELGTAGLQGLKQVSDSHRTATSSILSVHGSNVYPLSCTKY
IVGARPLLFPRLTSDVRDLKKKPLEFCKD"
7389..7940
/gene="Cpnl053"
7389..7940
/gene="Cpnl053"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19190.1"
/db_xref="GI:4377387"
/translation="MNFAKIDHNLHYITCLGIDIGVACPILSTDCILPNYSEKASHEVLY
YSKFCISGEPSRLATNPTYYISVSLPIGLRYEVTPSPGRHDFNIDMHVAPKIGAV
LSHGTRAKEIPGSKDYAFPSLTARESMLISEKLAMTPOVSEVIONCYSOCTVTKT
NLKEQYRHLSHNTGFLSVKSAF"
8270..10705
/gene="Cpnl054"
8270..10705
/gene="Cpnl054"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAD19191.1"
/db_xref="GI:4377388"
```

```
/translation="MQVLLSPQLPPPHQHSVGSISPSKRLVLAIATFLVFGMLLIISG
ALFLTIGLGAATISFGLIGLSALGQKIDLDQIDIEVACLRKLDKSKYSESLNDKREL
APSEPALQAQATLAQPKFELDQIDIEVACLRKLDKSKYSESLNDKREL
PFDVVEDTLISIFELRQIVAQEGWDLNGLNGRSLMNTAESLDFPHSKRLGYL
PSDVGEGGLKSAKEIVARLMSLHCEIKHVAFDRNSYMAEKAFKALGALAESV
YRSLTQSYRDKTLESEKAIIPWNGHITWLRDDAKSGCAKKKLRDAERWKKPKAVFW
VEEDGGFDNLLGDMGTVLDPIYQERMEITTHELYKTTFLKRLHRCALAKTTFE
KRSKNLQAEATPENYRCCVREQEDYKKEKREAFRRGNKILSPELESLSQPD
TRSNLKAQVATEPNYRCCVREQEDYKKEKREAFRRGNKILSPELESLSQPD
HGLNFESEKLEMEGLHILKLOKATVAVENKILSDAESLEIVFDVEMPORIEIE
KTLNMAELPLPTKAFKAGSOVNSCAEMLEKVPYCKESLAYVSKRSLYSLDEDL
RRAYTECOKRFQDGLSESEVAGRBOLRERIQEFETQGLDLYVEKLLVCVSRLANTE
CDCVSGYKKEAPPKRFTAQYIDELIYRVQSRNMTSRLREGVQACNKKMAKGLSE
EDKVLKEEYINLIRERKNKRLVGTIVATQORVAAVESIEVPEIPAPEKPSLL
DKARSFTREDHT"
BASE COUNT 3261 a 1883 c 2436 g 3267 t
ORIGIN
Query Match 100.0%; Score 1086; DB 1; Length 10847;
Best Local Similarity 100.0%; Pred. No. 4e-244;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGGCAAAAGTACCTCAAAACGAGAAAGATATTCTCTTGATTGTAAGGGGAATTGATTC 60
DB 826 TTGGCAAAAGTACCTCAAAACGAGAAAGATATTCTCTTGATTGTAAGGGGAATTGATTC 885
QY 61 TAACAGACCTTGGTCGAGCATCCATCTTAATTTGAAGTGAGTTCTACCACTTCATCAAA 120
DB 886 TAACAGACCTTGGTCGAGCATCCATCTTAATTTGAAGTGAGTTCTACCACTTCATCAAA 945
QY 121 ATGCTTATTCAAAATAAAGTCTCTTGTGGAGTTGATGATTGAAGGGAAGCAATAAT 180
DB 946 ATGCTTATTCAAAATAAAGTCTCTTGTGGAGTTGATGATTGAAGGGAAGCAATAAT 1005
QY 181 CTGATCCAAATTTAAAGGGAGAACACGTACGTTATCAATGAAGAGCGTGCCAAAGTTCCTG 240
DB 1006 CTGATCCAAATTTAAAGGGAGAACACGTACGTTATCAATGAAGAGCGTGCCAAAGTTCCTG 1065
QY 241 AGGAGAACTGATAAGAACGGCTCCATATGCTTTCTCTTCCTTCATGTTTCAATAAGTCC 300
DB 1066 AGGAGAACTGATAAGAACGGCTCCATATGCTTTCTCTTCCTTCATGTTTCAATAAGTCC 1125
QY 301 GCTTTCAACAGTAAACCAAGCAGGTACATAGCGATCAGATTGCTTTCTAGGGTTG 360
DB 1126 GCTTTCAACAGTAAACCAAGCAGGTACATAGCGATCAGATTGCTTTCTAGGGTTG 1185
QY 361 TATGGTGTGTTTTTACTAGGAAGAGCTGTACTTTTCTATGACTTTAGTGAAGAGTCT 420
DB 1186 TATGGTGTGTTTTTACTAGGAAGAGCTGTACTTTTCTATGACTTTAGTGAAGAGTCT 1245
QY 421 TCCCATGTTTATGAAAATAATCAGAAATGAGGATGAGAAGCCAGGACGTCGCCATA 480
DB 1246 TCCCATGTTTATGAAAATAATCAGAAATGAGGATGAGAAGCCAGGACGTCGCCATA 1305
QY 481 AAGGTCATGATGAGATCAGGATTAAGGAGAAATATCTTTATGAGAGATTCCGATTAC 540
DB 1306 AAGGTCATGATGAGATCAGGATTAAGGAGAAATATCTTTATGAGAGATTCCGATTAC 1365
QY 541 AGAGCAATGGGAAGTAACGGTCTCGCAACAGACAGACAGTAGTATTGATGGGGAGGTACAAA 600
DB 1366 AGAGCAATGGGAAGTAACGGTCTCGCAACAGACAGACAGTAGTATTGATGGGGAGGTACAAA 1425
QY 601 TCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAAATTTAATGACTGCTTG 660
DB 1426 TCCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAAGAAATTTAATGACTGCTTG 1485
QY 661 ATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAGTAATCAAGAAAAATCTTGG 720
DB 1486 ATGCTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAGTAATCAAGAAAAATCTTGG 1545
QY 721 ACAATAGCTTTTCCATAGACAAAACCGTGAGGAAGGAGTCCGATACCAATAATTCGATT 780
DB 1546 ACRATAGCTTTTCCATAGACAAAACCGTGAGGAAGGAGTCCGATACCAATAATTCGATT 1605
```

```
QY 781 AAAAAAGGAGTCTTACAAAGCCATATCAGTGGAGCAACCTTCCCTTGTGATCTTTGA 840
|||||
DB 1606 AAAAAAGGAGTCTTACAAAGCCATATCAGTGGAGCAACCTTCCCTTGTGATCTTTGA 1665
|||||
QY 841 TAATGACAGAAGATGAACAAAGAGTAAGTTTTTGAAGAAATCTTAAATAAGAAATGCAFTT 900
|||||
DB 1656 TAATGACAGAAGATGAACAAAGAGTAAGTTTTTGAAGAAATCTTAAATAAGAAATGCAFTT 1725
|||||
QY 901 GTGTCGAGCTAAAGCTTGCTCTCTTTATTTCTTTTCTTTGATGAAATGATTCGGTAGTAGGA 960
|||||
DB 1726 GTGTCGAGCTAAAGCTTGCTCTCTTTATTTCTTTTCTTTGATGAAATGATTCGGTAGTAGGA 1785
|||||
QY 961 ATATGGGCTCGAGTATGCACGTTGGAGTGATTTGGCTGTTCTGGAAGAACAGGAAAGTGA 1020
|||||
DB 1786 ATATGGGCTCGAGTATGCACGTTGGAGTGATTTGGCTGTTCTGGAAGAACAGGAAAGTGA 1845
|||||
QY 1021 TTGCTCAGCTTTAGAGCAATCTTCAGAAATATCTTTGGGTCCAGGTTCTCTCGCAGTA 1080
|||||
DB 1846 TTGCTCAGCTTTAGAGCAATCTTCAGAAATATCTTTGGGTCCAGGTTCTCTCGCAGTA 1905
|||||
QY 1081 GTGCAC 1086
|||||
DB 1906 GTGCAC 1911
|||||

RESULT 3
AE002240/c 22589 bp DNA linear BCT 30-MAY-2000
LOCUS Chlamydomophila pneumoniae AR39, section 68 of 94 of the complete
DEFINITION genome.
ACCESSION AE002240 AE002241 AE002161
VERSION AE002240.2 GI:8163502
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae AR39.
REFERENCE
1. (bases 1 to 22589)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE
1. (bases 1 to 22589)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
TITLE Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
JOURNAL Medical Center Dr, Rockville, MD 20850, USA
COMMENT On or before Jun 1, 2000 this sequence version replaced gi:7189730,
gi:7189720.
FEATURES
Source
1. .22589
Location/Qualifiers
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/notes="synonym: Chlamydia pneumoniae AR39"
complement(161..955)
/gene="Cp0805"
complement(161..955)
/gene="Cp0805"
/notes="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710
PID:216458; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
```

```
/product="dihydrodipicolinate reductase"
/protein_id="AAF38603.1"
/db_xref="GI:7189721"
/translation="MGSSMHVGVICSGRTGKIVSALEDSSEYVTLGPGFSRSSALTV
FYVTAHNDVLVDFSHPLTKKVVAHLILSPKPLIIGTGGPKCKEADHSLRELTTHV
FOVYCPNNSLGAYTHKELWMLISQLCNQFDIRRETHRRYKDKSLSGTAQDLDDTIQ
QVKADEGEXYECQDSKKTLEVOSSRGVDIPGEHEVAFISGGEOLLVRHTVFSRN
VFGKGLISLIDWLKTLNPNQGLYSGDTLELVLRNEHCCLKKTTDH"
832..1920
/gene="CP0806"
832..1920
/notes="biopterin-dependent aromatic amino acid hydroxylase
family protein; identified by match to PFAM protein family
HMM PF00351"
/codon_start=1
/transl_table=11
/product="biopterin-dependent aromatic amino acid
hydroxylase family protein"
/protein_id="AAF73705.1"
/db_xref="GI:8163503"
/translation="MHYCERTLDPKYILKALKRQSLSLFFQNSQSLQRAYSTPYSY
YRIILQENKEKQALARKHCTISILEFFKNLLEVLHLSLKNQREGCTDMVYVTPFF
NRNLWYLLSRLSKYKSPFFLDYLEAGLLSDFLDHQAVIKFFELTHFSYTPV
SGFVAPHOYLSLQDRYFPASVARTLDKDNFSITPDLIHDLGLHLPWLLHPSFSEFF
INMGRLFTKVIKVOALPKSKKQRTQTLQSNLIAIVRCFWFVESGLIENHEGRKAYGA
VLISSPQELGHAFIDNVRLPLLELDQIIRLPNTSTPQETLFSIRHDFELVELTSKLE
WMLDGLLESIPLYNQEKYLSGFEVLQ"
1911..2609
/gene="CP0807"
1911..2609
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38604.1"
/db_xref="GI:7189722"
/translation="MPMNTSHRKTLPVFSLSSTFTLLLVLSNLSKSLIPTTFNFPI
IPGAILPLPLTSLSDVNVNEIFGKARVMISAFIANLSSIVQIFNFFPVASPEM
OTAWGLDGLPLRFLASLLAFIVSQDLIDVLYTFKNTFNSSLMKSGSTWISQI
PTFIVDVCILYFGNGLSFPOTLINMFYSYIKITFCVLTTPFLYLAVENTIRKFLGMP
STKIANTVPLINQP"
2816..3811
/gene="CP0808"
2816..3811
/notes="similar to GB:J04423 SP:PI2996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF38605.1"
/db_xref="GI:7189723"
/translation="MREETVSNLSLELREIYHTPVFELTHKANAILRNSFLHSELOTG
YLISIKTGCVEDCAYCAOSSRYHTVTPPMKIVDVVERAKRAGELGATRVCLGAA
WRNAKDDRYEDRLVAMKSIDLCAEVCALGMSEQAQKLYDAGLYAYNHNLDSSP
EYFVTLITPESYEDRLTLDVNVKSGISCTCCGICVCGSEEDIKLHLVLAIRSHIP
ESYVNLITPESYEDRLTLDVNVKSGISCTCCGICVCGSEEDIKLHLVLAIRSHIP
LCELAGNSIFIGFKLLIVENNDDDAEMIKLGLIPRPSGIERGNFCYANNS"
3790..4938
/gene="CP0809"
3790..4938
/notes="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/notes="8-amino-7-oxononanoate synthase"
/protein_id="AAF38606.1"
/db_xref="GI:7189724"
```

/translation="MLCQQLFLIALARRKSKHTYRSLSINSHLIDFTSNDYLGFASSP
 ERKYYIKLHAIESLGATGSRLLTGHOSLCORIEQOLAAYHNPFESCLIFWTGYTANL
 GLYALATQDRIIDLDYIHASIVDGIIRLSKQSPFNHNDLNHLEKRLASHLGRTE
 VCVESVYSGSVAPLQAIASLCEKRYSAVLIIVDEAHAVGVFGDQEGVLSALGQDQV
 LTVTFKALGTGACALAGSSILKDYLLNFCRPFLYTTAOPPHALTAIELAYEHNOR
 AFNQRHLSALIHFRKAKNLGLOLMKONTTPTOSICVSGSHPARQAAQIQNSGY
 DVPRIYPTVQORELLRICLHAFTNKNEIDLHLHTLEQIFLCNVSSL"
 4917..5576
 /gene="CP0810"
 CDS
 4917..5576
 /note="similar to SP:P53558 PID:1277028 PID:2293186
 GB:AL009126; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthase B10d, putative"
 /protein_id="AAF38607.1"
 /db_xref="GI:7189725"
 /translation="MQRIIVGDTGVGKTIYSAILARALNAEYWKPIQAGNLSDS
 NIVHLSGAYCHPEAYRLHRPLSPHKAAGIDNYSIEESHICAPKTTSLNLIETSGGFL
 SPCTSKRQGVFSWSQWLLVSOAYLGSINHTCLTVEAMSRNLNLGVMVNGYPE
 DEEHLWTQEIKLPIIGTLAKEKEITKTIISCVAEQWKEVWTSNHQIGVSGTPELNL
 H"
 5513..6784
 /gene="CP0811"
 CDS
 5513..6784
 /note="similar to GB:J04423 SP:PI2995 PID:457106
 PID:490217 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="adenosylmethionine-8-amino-7-oxononanoate
 aminotransferase"
 /protein_id="AAF38608.1"
 /db_xref="GI:7189726"
 /translation="MDKQSSGNSGCIMHPFTQSALDSTPIKIVRGEGAYLVAESGTRY
 LDAISSWNCNLHGHPYIKLCEQAKLEHVIIFANFTHPEALEKSLAPLLEGL
 ERFPSDFSGTSIEMKIAVOYVYQNKAKSHFVGLSNAYHGDTPFGAMSTAGTSPT
 VFPHDLFSTIAAPYIGKEELAIAQATFVSESNIAAFIYEPILLOAGAGMLWYNE
 GKLEIKLAKGVLCIADBLTGFGRTGFLFASETDIPDIICLSKGLTGYPLIA
 LVTTKEIHDFAVSQDRMKALLHGTFTGNPLGCSAASLADLTLSPECLQOORMIER
 CHQFEQEAHGLWQRCVGLTVALDYPAEATGYFSQVDRHLNRFELRGVLLRPLGN
 TLYVLPPYCIQIEDLRIIYSHLODALCLOPQ"
 6766..7437
 /gene="CP0812"
 CDS
 6766..7437
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF38609.1"
 /db_xref="GI:7189727"
 /translation="MSTTIVKHFIHTASRWEPLKIEVASNVWHAQWINTLSPLNENG
 AKKISASHPTEVKEEVKHAEEFRHGHYKLTQISRISETSLPDYTSKNLLGLLTK
 YTHLLDLTRCVLENSYSLSGOTLKTAAYILVYAIELRASELYPLHYDLILKEQSK
 ITVKSIILEEOGHLOEMERELKDLPHGEELLGYACQFEGELCLQFVERLEQMIFDPS
 TTFKR"
 7508..8845
 /gene="CP0813"
 CDS
 7508..8845
 /note="3-phosphoshikimate 1-carboxyvinyltransferase;
 identified by match to PFAM protein family HMM PF00275"
 /codon_start=1
 /transl_table=11
 /product="3-phosphoshikimate 1-carboxyvinyltransferase"
 /protein_id="AAF37306.1"
 /db_xref="GI:8163504"
 /translation="MLYKYSPSSVYGNAPIPSSKSHTLRAILHSAVECKSIYNYL
 DSPDPEAMICKOMGASIKKFPQILEIVGNFLAIFPKYTLIDAGNSGIVLRFMTALA

CVFSKEITVTGSQLQRRMAPLQALRNFGASFHSSDKSVLPFTTMSGLPSRAYS DV
 EGSDFPAGALAVACSLAFSGPCSTFIIEPKRPMFDSLAWLEKHLHPYSCDSTVYSP
 PGSSHQFSGSYHVTGDFSSAATIAAALLSKSLQIRLNLIDLDQDKIFSLMON
 LGASIOYDNEELVPPSPFSGSDIMDCIDALPILVLCPPADSPSHLYNARSADK
 ESRIILATTEELQKMGACIQPHDGLVNPSPLYGAVLSDHDDHRLAMALTAAALYAS
 GDSRIHNTACVRKTPFPNFVOTLINMEARIECHDNYSMWSTHKKRYFARESFG"
 8781..9311
 /gene="CP0814"
 CDS
 8781..9311
 Query Match 100.0%; Score 1086; DB 1; Length 22589;
 Best Local Similarity 100.0%; Pred. No. 3.7e-244;
 Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGGCAAGTAGTACCTCAAAACCCAGAAAGATATTTCTGTGATTTGTAAGGGGAATTGATTC 60
 DB 1917 TTGGCAAGTAGTACCTCAAAACCCAGAAAGATATTTCTGTGATTTGTAAGGGGAATTGATTC 1858
 QY 61 TAACAGACCTTTGGTCGAGCATCCATTTCTAATTTTGAAGTGAGTTCTACCACTTTCATCAAA 120
 DB 1857 TAACAGACCTTTGGTCGAGCATCCATTTCTAATTTTGAAGTGAGTTCTACCACTTTCATCAAA 1798
 QY 121 ATGTCTTTATGAAATAAAGTCTCTTGTGGAGTTGATGTATTGAAGGAAGACCAATAAT 180
 DB 1797 ATGTCTTTATGAAATAAAGTCTCTTGTGGAGTTGATGTATTGAAGGAAGACCAATAAT 1738
 QY 181 CTGATCCAAATTTAAAGGGAGACACGTAGCTTATCAATGAAGCGTGTCCAACTTCCTG 240
 DB 1737 CTGATCCAAATTTAAAGGGAGACACGTAGCTTATCAATGAAGCGTGTCCAACTTCCTG 1678
 QY 241 AGGAACTGATAGAAGCGGTCCATATGCTTTTCTCTCATGCTTTTCAATAAGTCC 300
 DB 1677 AGGAACTGATAGAAGCGGTCCATATGCTTTTCTCTCATGCTTTTCAATAAGTCC 1618
 QY 301 GCTTTCAACAGTAAACCAAGCAGCGTACAATAGCGATCAGATTGCTTTGTAGGGTTG 360
 DB 1617 GCTTTCAACAGTAAACCAAGCAGCGTACAATAGCGATCAGATTGCTTTGTAGGGTTG 1558
 QY 361 TATGCTTTGTTTTTACTAGSAGAGAGCTGTACTTTTTCTATGACTTTTAGTGAAGAGTCT 420
 DB 1557 TATGCTTTGTTTTTACTAGSAGAGAGCTGTACTTTTTCTATGACTTTTAGTGAAGAGTCT 1498
 QY 421 TCCCATGTTTATGAAAATTCAGAAAATGAGGATGATAGAAGCAAGGACGTCGCCATA 480
 DB 1497 TCCCATGTTTATGAAAATTCAGAAAATGAGGATGATAGAAGCAAGGACGTCGCCATA 1438
 QY 481 AAGGTCATGGATGAGATCAGAGTTAAGGAGAAATTTATCTTTATCGAGAGTTCCGATTAC 540
 DB 1437 AAGGTCATGGATGAGATCAGAGTTAAGGAGAAATTTATCTTTATCGAGAGTTCCGATTAC 1378
 QY 541 ACAGCAATGGAAAGTAACGGTCTGCAACAGACAGATATTGATGGGAGCTACAAA 600
 DB 1377 ACAGCAATGGAAAGTAACGGTCTGCAACAGACAGATATTGATGGGAGCTACAAA 1318
 QY 601 TCCGAAACGGGATAATAGGAAAATGTGTTCTTAATTCGAAAGATTTAATGACTGCTTG 660
 DB 1317 TCCGAAACGGGATAATAGGAAAATGTGTTCTTAATTCGAAAGATTTAATGACTGCTTG 1258
 QY 661 ATGGTCTAAGAAATCAGAAAGGAGACCCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGG 720
 DB 1257 ATGGTCTAAGAAATCAGAAAGGAGACCCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGG 1198
 QY 721 ACAATAGCTTTTCCATAGAGAAAACCGTGAGAAAGAGTTCGATACCAATAATTCGGATT 780
 DB 1197 ACAATAGCTTTTCCATAGAGAAAACCGTGAGAAAGAGTTCGATACCAATAATTCGGATT 1138
 QY 781 AAAAAGGAGTGTCTACAACGCCATATCAGTGGAGCAACCTCCCTTCATCTTTTGA 840
 DB 1137 AAAAAGGAGTGTCTACAACGCCATATCAGTGGAGCAACCTCCCTTCATCTTTTGA 1078
 QY 841 TAATCAGAGATGACAAGAGTAGTTTGTGAATTTCTAATAATAGAAATGCAATTT 900
 DB 1077 TAATCAGAGATGACAAGAGTAGTTTGTGAATTTCTAATAATAGAAATGCAATTT 1018

EKDRIFRTHTHDIEMKEIGFCCKGIENYSRHFTGAPGAPPTCLDYFPEDFLIIDE
SHOTLPQIRAMRDQSRKSLVEYGRFLPSADNRDLTLEAEQYRKRYVYSATPG
DTEVESSGHQVQIIRPTQIPDMPPEIRPAGQVDVLDLEIRLSQKHEKILIVISI
TKKLADWAGLSELEIPAAYSLSHGETAERTQILTDLSRGVIDLVGLNLLREGDL
PEVSLVAILDADKCGFLRSSTSIQFCGGAARNKNGKVFYADQKRSIETRETER
RQIQOLDYNKHNTVPKPIIKALFANPILQTSKDSKESPKSORPLSLEDLEEQIKYE
ALMQRAAEFRFNAAKYRDMAQCKEQLLYLF"
complement(7327. .8361)
/gene="trps"
complement(7327. .8361)
/gene="trps"
/codon_start=1
/transl_table=1
/product="tryptophanyl tRNA synthetase"
/protein_id="BAA99010.1"
/db_xref="GI:8979176"
/translation="MNKKRVRLGDRPTGKHLGHVWSIKNRLSELNPEYDCFFII
ADHLTKIKESVLDVLDWLSVGDIDPKTSIYVLSASPIEYELHLDFS
ALISNRMVGIPLSKMARNASIEGSLSYGLIGYPILOSADILLAKAQFVYVKGONE
HIVETLRDIARNRLXQGVPEPEVLOGELTSLVIGDQGMKSKANNAYILSDA
TITERKRMVTDPNRIATTPGRVGNPLFIYHDFNPHKDEVEFRKARYQGCIDKI
EVKARLAEELHFKPIKRSERSEFLSKPLALQNVLEDGTHKMRVAKVTMEVHDKFG
FSHKWRSLLK"
complement(8370. .8924)
/gene="CPJ0803"
complement(8370. .8924)
/gene="CPJ0803"
/codon_start=1
/transl_table=1
/product="CT584 hypothetical protein"
/protein_id="BAA99011.1"
/db_xref="GI:8979177"
/translation="MAAKTKLEBNDVLLLEGNLKRIFATPIGYTTFRFQNVNPN
CANGQOEIAFFEMLINGLTQELAPQQAHLIAEFMMPIRKARDTHERGEFIN
FITSMDLTQQRCEIFLARLAVDGOEFLMTDVQNTCHLIRLLARLLEAKNPVGEK
NLQEOEETSLKNHFDLTKALO"
complement(8928. .9680)
/gene="gp60"
complement(8928. .9680)
/gene="gp60"
/codon_start=1
/transl_table=1
/product="CHLTR plasmid paralog"
/protein_id="BAA99012.1"
/db_xref="GI:8979178"
/translation="MGNLKTLLSRRFKKNTPTKMEALARKWEGDPSPLAVRLSNPTL
SSKEKQRLHLLHONYNREQIEPDLTOLCTLSAPVKOIHOSVLLHGERITKVRDL
KSYRCAGASSLLITYGNROTPTNFVYVELFTLLPELPKLEMEKMFQAVYLLASRQ
RGAQKEEIIIRNRRKSELDRIKREFPLVETDCRTSPVKQLAMLTNGSOLTK
CISLSDSQIIEKLKLEKVKSNLFPDTKV"
complement(9685. .10452)
/gene="minb"
complement(9685. .10452)
/gene="minb"
/codon_start=1
/transl_table=1
/product="chromosome partitioning ATPase"
/protein_id="BAA99013.1"
/db_xref="GI:8979179"
/translation="MKTIAYNSFKGTAKTSTTLHLGAALQYHOARVLLIDPDAON
LTSGLGLDPCYDLAVVLOGEIOEVIRIQDTOLDLIPADTWLERIYVSGNLAD
RYSHERLYVLGSDQVDYVIDIITPESCLWLTESALIAADYALICATPEYVSKGLE
RGAFTGQISARHPLTLILGVALSFWNCRCNKNFAELIHKHTFPGKLLNTKIRDIIV
SPAAIHGKPVATSPSARASEDYENLTKEILLIRDI"
complement(10485. .12392)
/gene="thrs"
complement(10485. .12392)
/gene="thrs"
/codon_start=1
/transl_table=1
/product="threonyl tRNA synthetase"
/protein_id="BAA99014.1"
/db_xref="GI:8979180"

/translation="MIQVTCQKNYEVLEGTAAELAKQLKNSHQFIGVLINERPRDL
Query Match 100.0%; Score 1086; DB 1; Length 325865;
Best Local Similarity 100.0%; Pred. No. 2.7e-244;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGCAAGTACCTCAAAACGAGAAAGATATTCTCTGATTGTAAAGGGAATTCATTG 60
DB 295293 TTGGCAAGTACCTCAAAACGAGAAAGATATTCTCTGATTGTAAAGGGAATTCATTG 295352
QY 61 TAACAGACCTTGGTCGAGATCCATCTTAATTTTGAAGTGGTCTACCAAGTTTCATCAA 120
DB 295353 TACAGACCTTGGTCGAGATCCATCTTAATTTTGAAGTGGTCTACCAAGTTTCATCAA 295412
QY 121 ATGCTTATTAATAAAGTCTCTTGTGGAGTTCATCTATTGAAGGGAAGCAATAAT 180
DB 295413 ATGCTTATTAATAAAGTCTCTTGTGGAGTTCATCTATTGAAGGGAAGCAATAAT 295472
QY 181 CTGATCCAAATCTAAAGGGGAGAACACGTACGTATTCAATGAAGCGTGTCCAAAGTTCCTG 240
DB 295473 CTGATCCAAATCTAAAGGGGAGAACACGTACGTATTCAATGAAGCGTGTCCAAAGTTCCTG 295532
QY 241 AGGAACTGATAGAAGCGCTCCATATGCTTTTCTCTCTTCATGCTTTTCAATAAGTCC 300
DB 295533 AGGAACTGATAGAAGCGCTCCATATGCTTTTCTCTCTTCATGCTTTTCAATAAGTCC 295592
QY 301 GCTTTCACACTAAACCAAGCGGTACATAGCGATGAGATGCTTTCTAGGGTTTG 360
DB 295593 GCTTTCACACTAAACCAAGCGGTACATAGCGATGAGATGCTTTCTAGGGTTTG 295652
QY 361 TATGCGTTGTTTTTACTAGGAAGAGCTTGACTTTTCTATGACTTTAGTGAAGAGTCT 420
DB 295653 TATGCGTTGTTTTTACTAGGAAGAGCTTGACTTTTCTATGACTTTAGTGAAGAGTCT 295712
QY 421 TCCCATGTTTATGAANAATTCAGANAATGAGGATGTAGAAGCCAAAGCACGTGCCCTAA 480
DB 295713 TCCCATGTTTATGAANAATTCAGANAATGAGGATGTAGAAGCCAAAGCACGTGCCCTAA 295772
QY 481 AAGGTCTATGATGAGATCAGGAGTAAAGGAGAAATTAATCTTTATCGAGAGTTCGCATTAC 540
DB 295773 AAGGTCTATGATGAGATCAGGAGTAAAGGAGAAATTAATCTTTATCGAGAGTTCGCATTAC 295832
QY 541 AGAGCAATGGGAAGTAAGCGTCTGCAACAGACAGCAAGTATTGATGGGAGCTACAAA 600
DB 295833 AGAGCAATGGGAAGTAAGCGTCTGCAACAGACAGCAAGTATTGATGGGAGCTACAAA 295892
QY 601 TCCGCAACCGGATATAGGAANAATGTGTTCTTAATTCGAAGAATTTAATGACTGCTTG 660
DB 295893 TCCGCAACCGGATATAGGAANAATGTGTTCTTAATTCGAAGAATTTAATGACTGCTTG 295952
QY 661 ATGGTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAAAATCTTGG 720
DB 295953 ATGGTCTAAGAAATCAGAAAGGAGACCGGAAAGCTTCTTAAGTAATCAAGAAAAATCTTGG 296012
QY 721 ACAATAGCTTTTCCATAGAGAAACCGTGGAGAAAGAGTCCGATACCATTAATTCGGATT 780
DB 296013 ACAATAGCTTTTCCATAGAGAAACCGTGGAGAAAGAGTCCGATACCATTAATTCGGATT 296072
QY 781 AAAAAGGAGTCTTCAACACGCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTTGA 840
DB 296073 AAAAAGGAGTCTTCAACACGCCATATCAGTGGAGCAACCTTCCCTTTGATTCTTTGA 296132
QY 841 TAATCAGACAGATGACAAAGAGTAGTAACTTTTGAANAATTTCAANAATCAANAATGCATTT 900
DB 296133 TAATCAGACAGATGACAAAGAGTAGTAACTTTTGAANAATTTCAANAATCAANAATGCATTT 296192
QY 901 GTGTCGAGCTAAAGCTTGCTTCTTTATTTTCTTTTGTGAGATGATTCGGTAGTAGGA 960
DB 296193 GTGTCGAGCTAAAGCTTGCTTCTTTATTTTCTTTTGTGAGATGATTCGGTAGTAGGA 296252
QY 961 ATATGGGCTCGAGTATCGACGTTGGAGTGAATTTGGCTGTCTGGAAGAACAGGAAAGTGA 1020
DB 296253 ATATGGGCTCGAGTATCGACGTTGGAGTGAATTTGGCTGTCTGGAAGAACAGGAAAGTGA 296312

* 4931 5030: gap of 100 bp
 * 5031 85903: contig of 80873 bp in length
 * 85904 86003: gap of 100 bp
 * 86004 97976: contig of 11973 bp in length
 * 97977 98076: gap of 100 bp
 * 98077 147748: contig of 49672 bp in length
 * 147749 147848: gap of 100 bp
 * 147849 151120: contig of 3272 bp in length.

FEATURES

source
 1..151120
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RP11-21M15"
 /clone_lib="RPC1-11 Human Male BAC"
 1..4930
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 5031..85903
 /note="assembly_fragment"
 86004..97976
 /note="assembly_fragment"
 98077..147748
 /note="assembly_fragment"
 147849..151120
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 47047 a 27918 c 27085 g 48670 t 400 others

BASE COUNT
 ORIGIN

Query Match 4.3%; Score 46.8; DB 2; Length 151120;
 Best Local Similarity 47.9%; Pred. No. 0.8; Mismatches 147; Indels 0; Gaps 0;
 Matches 135; Conservative 0;
 Oy 598 AATCTCGAAGCGGATATAGGAAAAATGTCTTAATTCGAAGATTTTAACTGCTGC 657
 Db 13385 AATCTCAAAATAGTACAGGAGAAAAATACGAAGATCAAGCAAAAAATAACAAAAATA 13444
 Oy 658 TTGATGCTCTAAGAAATCAGAAGGAGAGACCCGAAAGCTCTTAAGTAATCAAGAAAAATCT 717
 Db 13445 GAGATTAAAAAATAAAGATCAATGAACAGAGTTTTTCTTTTAAAAAATAAGTAAAAAT 13504
 Oy 718 TGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAGGAGCGATACCAATAAATTCG 777
 Db 13505 TGACACATCTTTTAGCTACACTAGAAAAAAGAGAGAGAGACTCAAAATTTACAAAAATTAA 13564
 Oy 778 ATTAATAAAGGGAGTCTTACACAGCCATATCAGTCGACACCTTCCTTTGATTCTT 837
 Db 13565 AGGTGAAAAGGNAACATTAATAACTGACACCACTACACAGTATACACAGACTAT 13624
 Oy 838 TGATATGACAGAGATGACAAAGAGTAGTATTTTGAAGAAA 879
 Db 13625 TATGACAACTGTATACCAACAAATTTGCAAAATCTAGAAAA 13666

RESULT 7

AC084882/c
 LOCUS
 DEFINITION Homo sapiens chromosome 15 clone RP11-408J6 map 15q21.1, complete sequence.
 ACCESSION AC084882 AC019253
 VERSION AC084882.2 GI:13430989
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS
 1 (bases 1 to 193804)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
 Nesbitt,R., Traicoff,R. and Hood,L.
 TITLE
 Sequencing of human chromosome 15 D15S117 region

JOURNAL
REFERENCE

2 (bases 1 to 193804)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
 Nesbitt,R., Traicoff,R. and Hood,L.

TITLE
JOURNAL

Direct Submission
 Submitted (28-NOV-2000) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

REFERENCE
AUTHORS

3 (bases 1 to 193804)
 Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
 Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., Kaur,A., Madan,A.,
 Nesbitt,R., Traicoff,R. and Hood,L.

TITLE
JOURNAL

Direct Submission
 Submitted (22-MAR-2001) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

COMMENT

On Mar 22, 2001 this sequence version replaced gi:11386266.

----- Genome Center

Center: Multimegabase Sequencing Center

Center code: UWMSC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowen@systemsbiology.org

Drafting center: WIBR

----- Summary Statistics

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-Primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

Note: Data from AC066615 [Drafting center: UWMSC], AC064813
 [Drafting center: WIBR], AC025843 [Drafting center: WIBR] and
 AC023905 [Drafting center: UWMSC] were added for finishing

FEATURES

source

1..193804
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="15"
 /map="15q21.1"
 /clone="RP11-408J6"
 /clone_lib="RPC1 human BAC library 11"
 /note="This clone overlaps RP11-90J19 AC066615,
 RP11-691D11 AC064813, RP11-313A9 AC025843 and CTD-2150N2
 AC023905. Data from overlapping BACs were added and the
 consensus sequence was determined from RP11-408J6 to the
 extent possible"

misc_feature

1..79521
 /note="overlap with RP11-90J19, AC066615"

misc_feature

1..74417
 /note="overlap with RP11-691D11 AC064813"

unsure

103950..104040
 /note="low quality data"

unsure

113560..113700
 /note="low quality data"

misc_feature

140705..193804
 /note="overlap with RP11-313A9, AC025843"

unsure

152960..152980
 /note="low quality data"

unsure

164005..164040
 /note="low quality data"

misc_feature

177480..193804
 /note="overlap with CTD-2050N2, AC023905"

BASE COUNT

57860 a 35778 c 37171 g 62995 t

ORIGIN

Query Match 4.3%; Score 46.8; DB 9; Length 193804;

Best Local Similarity 47.9%; Pred. No. 0.77; 147; Indels 0; Gaps 0;
 Matches 135; Conservative 0; Mismatches 147;

Oy 598 AATCTCGAAGCGGATATAGGAAAAATGTCTTAATTCGAAGATTTTAACTGCTGC 657

Db 104612 AATCTCAAAATAGTACAGGAGAAAAATAGCAAGATCAAGCAAAAAATAACAAAAATA 104553

```

QY 658 TTGATGCTCAAGAAATCAGAAAGGAGACCGGAAAGCTTCTAAGTAATCAAGAAATAATCT 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104552 GAGATTAATAAATAAAGATCAATCAACACAGAGTTTCTCTTTTMAAATAAATAGTAAAT 104493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 718 TGGACATACTTTTCCATAGAGAAACCGTGAGGAAGAGAGTCGATACCAATAATCCG 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104492 TGACACATCTTTTAGCTACACTAAGAAAAAAGAGACAGACTCAAAATTTACAAAAATTA 104433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 ATTTAAAAAGGAGTCTTTACACAGCCATATCATGTGGACACCTTCCCTTTTGATTCTT 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104432 AGGTGAAGAAGGAAATCAACATTAACCTGACACACAGTAATACACAGATCACACAGACTAT 104373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 TGATAATGACAGAGATGACAAAGAGTAAGTTTGTGAAAAA 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104372 TATGAACAACCTGTATACCAACAATTTGGAATAATCTAGAAAA 104331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AF135186/c 1642 bp mRNA linear INV 13-MAY-1999
LOCUS Caenorhabditis elegans tryptophan hydroxylase (tph-1) mRNA,
DEFINITION complete cds.
ACCESSION AF135186
VERSION AF135186.1 GI:4809152
KEYWORDS Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 1642)
AUTHORS Size, J. Y. and Ruvkun, G.
TITLE tph-1 encodes a C. elegans tryptophan hydroxylase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1642)
AUTHORS Size, J. Y. and Ruvkun, G.
TITLE Direct Submission
JOURNAL Submitted (16-MAR-1999) Department of Molecular Biology/Department
of Genetics, Massachusetts General Hospital/Harvard Medical School,
Boston, MA 02114, USA
FEATURES
source Location/Qualifiers
1..1642
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="II"
/gene="tph-1"
/cds="tph-1"
44..1642
/note="tph-1"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAD30115.1"
/db_xref="GI:4809153"
/translation="MDSLQFQASAMKFOYYSKKAAGTMSNVSMSDNRMDEPKRFR
RRSGLDGIPVPEVDYKQIFPTRTVRRASIREGDEEGVQILTIIVKSSRVSDIS
KMIANLDPRIKHELRDSODGSKTMDVLELPHYKQKQAMDLMRLNLDVHEV
SSTIRPTAIKEQYEPGSDATTGSEWPKSIYDLIDICAKRVIMYGAGLDADHPGKFD
TEYRHRMCPAEALANIKHGEPIRTEYTSSEKRWGIIYIKLRELHKHAKQKFLDN
FELLERHCYGSNNIPQLEDICKFLKAKTGFVRPVAQYLSARDFLAGLAYRVFFCTQ
YVRHHDPFTYPPEDTVHELMGHMALFADPDFQAFQSEIGLSLASEEDLKKLATLY
FFSLEFGLSDDAADSPVKENGSNHREFKVVYAGLLSAGELQHAVEGSAITIRFDPD
RVFQECILTTDFQSAFYFRNFEEAQOKLRMTNNMRPFIVRNPYTESVLENNRS
SIMLVNLSRLSDNLLAGALHVL"
BASE COUNT 479 a 347 c 351 g 465 t
ORIGIN
Query Match 4.3%; Score 46.6; DB 3; Length 1642;
Best Local Similarity 45.5%; Pred. No. 1.5;
Matches 249; Conservative 0; Mismatches 289; Indels 9; Gaps 2;
QY 113 TCATCAAAATGCTTATTGAAATAAAGTCTCTTGTGGAGTTGATGATTTGAAGGAAGA 172
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1464 TCCTTCAAAATTTCTAGTATAGAAATACGCTGACTGGAAGTAGTAATGAGACATCTTTC 1405
QY 173 CGAATAATCTGATCCAAATTCATAAGGAGAGAACACGCTAGCTTATCAATGAAGCGTGTCCA 232
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1404 TCACACACAGATCCGGATCAAAACGAATAATGTTGCACTACCCCTCAACGCAATGTGC 1345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 233 AGTTCTTGAGGAGAACTGATAGAACCGCTCCATATGCTTTCTTCTTCATGTTTTCATCA 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1344 AACTGCCACGACTGCTCAGAAAGTCTGCTCCGTATACTTTAAATCTTTTCATGATTTGAT 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 293 ATAAGTCCGCTTTCACAGTAAACCAAAAGCGGTACAAATAGCGATCAGATGCTTTTGT 352
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1284 CCATTCTTTTACTGGAG-AATCGGACGCTCATCAGAGAGACCAAAATCAATGGA 1226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 353 AGGTTTGTATGCTTTGTTTTTACTAGGAAGAGCTTTGTACTTTTTCTATGACTTTAGTG 412
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 AAACAAGTAGAGTGTTCGAAGCTTCTTCAAA-----TCTTCCTCTGATGCTCCAAGA 1174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 413 AAGAGTCTTCCCATGTTTATGAAAAATTCAGAAAATGAGGATGTAGAACCAAGGCACG 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1173 GAAGCTAATCAATCTCTTGAGAAAACTGACAAAATCTGATCAGCGAATAGAGCCATG 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 473 TGCCCTAAAAGGTCATGATGATGAGATCAGAGTAAAGAGAAATATCTTTTATCGAGAGTT 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1113 TGACCATGAGCTCGTGAACGGTGTCTGTTCTGGAGTGTAAAATGATCGCATGATGG 1054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 533 CGCATACAGAGCAATGGAAGTAAGTAAAGTAAAGTCTGCAACAGACACAAGTATTGATGGGA 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1053 CGAAGCTATTGATGTCAGAGAAGACACACATATGCAAGACCTCCCAAGAAATCAGAGCT 994
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 GCTCAAAATCTGAAACGGGATPAATAGGAAAAATGTGTTCTAATTCGAAGAAFTTAATG 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 993 GATAAGTATCCGGCGACTGGCGACACACGGAATCCAGTTTGTCTTCAAAAACCTTGCAG 934
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 653 ACTGCTT 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 933 ATATCTT 927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
AF119388/c 1531 bp mRNA linear INV 23-MAY-1999
LOCUS Caenorhabditis elegans phenylalanine hydroxylase (K08F8.4) mRNA,
DEFINITION complete cds.
ACCESSION AF119388
VERSION AF119388.1 GI:4883766
KEYWORDS Caenorhabditis elegans.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 1531)
AUTHORS Loer, C. M., Davidson, B., and McKerrow, J.
TITLE A phenylalanine hydroxylase gene from the nematode Caenorhabditis
elegans is expressed in the hypodermis
JOURNAL J. Neurogenet. (1999) In press
REFERENCE 2 (bases 1 to 1531)
AUTHORS Loer, C. M.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Dept. of Biology, University of San Diego,
5998 Alcalá Park, San Diego, CA 92110, USA
FEATURES
source Location/Qualifiers
1..1531
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="II"
/clone="cosmid K08F8"
/dev_stage="mixed"
1..1531
/gene="K08F8.4"
/note="K08F8.4/bas-2"
90..1463

```



```

cDNA EST yk273b10.5 comes from this gene"
/codon_start=1
/protein_id="CAA91284.1"
/db_xref="GI:3878361"
/db_xref="SPTREMBL:Q21361"
/translation="MDVDSASLSPSSKSLVVCASAETSSSSSSSNFSGSPPEERR
NSVNESVIKDEHYWERRRRNDASRRRKRQNDLAMEEKIMLLSNAENKLSQGLT
TIPOSVTEPTSLIIPQAKNLFAPAGTIASQSSMLTVPLOASHIPSMQLQCO
LOPTTIVSTASTQOASTSASLFSSSSAFHPFSPSAOQSPSSSVLVIKER
RSPDSTVNNFQLOPGSSVTOIQGPAPSTPPVIAQVQOQSPILLSALKSQRRP
SPVTPQSTREIHTISGNSPRHTGNKSDCESVSSASFSPSHSDHSNYSNKSPOYVD
RRRNNEAKRCRANRAVFEYRSRRVOLLEGENEDLRITETLKAETAFHFKSVLAOR
ASVVTALHP"
complement(join(18851..19061,19149..19267,19316..19421,
19473..19561,19609..19713,19799..19939,19989..20176,
20245..20482,20530..20599,20934..21022))
/gene="K08F8.3"
complement(join(18851..19061,19149..19267,19316..19421,
19473..19561,19609..19713,19799..19939,19989..20176,
20245..20482,20530..20599,20934..21022))
/gene="K08F8.3"
/notes="contains similarity to pfam domain: PF00852
(Pucosyl transferase), score=875.1, E-value=6.9e-260, N=1
cDNA EST yk78d11.3 comes from this gene
cDNA EST yk78d11.5 comes from this gene
cDNA EST yk340d5.3 comes from this gene
cDNA EST yk340d5.5 comes from this gene"
/codon_start=1
/protein_id="CAA91285.1"
/db_xref="GI:3878362"
/db_xref="SPTREMBL:Q21362"
/translation="MKKQNTPEVGYATSCRMWGVDDRLPLWKYLMFACCTYLL
VIVAPTSQKQDKEGELEJNSDHELDVPILOKEELKPOQSPSEENVPKRKTFNEN
PVKEPFDVEVLTSQDKLEERTATVPIQKRLILSNAGNSQDNLOQGDQWNCQ
TQVRARAPADAVLIAHMDNDVPKNOYVYVFSQSPANSQIQRPRDIINMILGR
HDTPASPGTYTVKLGASRKRTQGVQDANLVNGKAKGAAMFVSHQCOTSKREDFVKKL
QHLIDIVGGCGPKARGSKDCLMDTDIDHYFTFENSICEDYVTEKLMKSGYQN
TLPIDVLVLRKEPPEVPPNSFIADDFKSKVMGDYLNKMNNTKATYEFYFWRIDYK
VVFDDGSHHDVLERPWGFCQVCRMAWTEPRQKVLIPNWDAYWRQTCEDGLTVDSIPL
D"
complement(join(24990..25164,25210..25439,25484..25610,
25786..25921,25972..26168,26281..26443,26496..26688,
27097..27213,27309..27344))
/gene="K08F8.4"
complement(join(24990..25164,25210..25439,25484..25610,
25786..25921,25972..26168,26281..26443,26496..26688,
27097..27213,27309..27344))
/gene="K08F8.4"
/notes="contains similarity to pfam domain: PF00351
(Riopterin-dependent aromatic amino acid hydroxylase),
Score=800.0, E-value=1.6e-252, N=1; PF01842 (ACT domain),
Score=19.5, E-value=2.2e-11, N=1
cDNA EST yk432h11.3 comes from this gene
cDNA EST yk432h11.5 comes from this gene
cDNA EST yk349f13.5 comes from this gene"
/codon_start=1
/protein_id="CAA91286.1"
/db_xref="GI:3979833"
/db_xref="SWISS-PROT:P90925"
/translation="MPPAGODDLDFKYMAYSVADNVNADIGKTIIVFTLREKAGALA
ETLKLQAHVDNLSHIESPSKTEGCEVILVEFAEADHRIKGVIEHFQOKAEKV
LVQWNTNKNQKDSVPWFQKINDIDQFANRLYSYGAELDADHPGKDMTYRERKK
FADIAFNFKGDKTPTITTYDEIATVTRNLTVPYKNACQEFNYIFPLQKNCQ
FGPDRIPQLOQVSDFLKDCGTGYIRFVAGLLSRRDFLAGLAFRVHSTQYRIHHSAPK
YTPEDICHELLHVLPLADVEFAQFSQIGLASLAPDDVIEKLATLWYFTFEGIC
QODGKKYAGAGLSSEGELOALSDKPEVDFPVCVCTKYPITEYQPKYFLAES
ASAKNKLKSWAATINRPFQIRYNAYTORVEILDVKAALQRLARDINDISTLEEALGK
VNNLKNK"
join(29693..29728,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5b"
join(29693..29728,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)

```

```

/gene="K08F8.5b"
/codon_start=1
/protein_id="CAA91288.1"
/db_xref="GI:3878365"
/db_xref="SPTREMBL:P90926"
/translation="MGRANDSGVVKCMFVLDKIVSGADVGGSTOLLILDLHIEKSKS
EYDWNVKKQSQSTISGKYEMVRCYSASMTWPEHGLLCVNTKEKLVAHFVDNEMOI
KTYDKLEULLQCKTAGRDATKEKEKSGTSKSLTIVNMDENFNGSAMPNKN
VNSDKNTPSAGIDQFVALYRHKPTMGRANOSNGRIEASFVDAKREFTCATVGSLO
LISLPPTYGDDYVWYTOADYDKQYVSHLSYICPAIVPVCDFKFLGOVNMKT
ECATVALNGSVTOIDGSEPAKSLVLCRKMAETMLI"
join(29858..29977,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5a"
join(29858..29977,30107..30376,30419..30588,30762..30896,
30946..31051,31100..31190,31450..31580)
/gene="K08F8.5a"
/codon_start=1
/protein_id="CAA91287.1"
/db_xref="GI:3878364"
/db_xref="SPTREMBL:Q21365"
/translation="MESHEVLIPEWPNIPLOGILPTVDMCSGKLDRAKFAQLICMFV
LDDKIYSGADVGGSTQLLILDLHIEKSKSFHYDWNFKKQASLTIEGKVMRCYSAS
LDDKIYSGADVGGSTQLLILDLHIEKSKSFHYDWNFKKQASLTIEGKVMRCYSAS

Query Match      4.2%   Score 45.6; DB 3; Length 41246;
Best Local Similarity 59.1%; Pred. No. 1.8;
Matches 78; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 191 TCTAAGGAGAACACGTCATCAATGAAAGCGTGTCACAGTCCTCGAGGAGAACTG 250
DB 25306 TCAARAATCAACATCTCGGCTTATCACTCAACGCATATGAAGCTCTCCAAATGAATC 25365

QY 251 ATAAGAACGGTCATATGCTTTCTTCCTCATGTTTCAATAAGTCGCTTTCAACA 310
DB 25366 AAGATCGGCTCGTAGCTTTTCTCCCATCTTGTGACAGATTCCAAAATTCGATT 25425

QY 311 GTAAACCAAAAG 322
DB 25426 GTGAACCAAGTAG 25437

RESULT 11
AC121238/c 109206 bp DNA linear HTG 23-AUG-2002
LOCUS Medicago truncatula clone mth2-34p9, WORKING DRAFT SEQUENCE, 5
DEFINITION ordered pieces.
AC121238
VERSION AC121238.8 GI:22218484
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 109206)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
JOURNAL Medicago truncatula BAC Clone mth2-34p9
REFERENCE 2 (bases 1 to 109206)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
JOURNAL Direct Submission
REFERENCE 3 (bases 1 to 109206)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
JOURNAL Direct Submission
REFERENCE 4 (16-MAY-2002) Department of Chemistry And Biochemistry,
JOURNAL the University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 109206)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
JOURNAL Direct Submission
REFERENCE 6 (23-AUG-2002) Department of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

```


OK 73019, USA
 On Aug 14, 2002 this sequence version replaced gi:22038527.
 ----- Genome Center
 Center: Department of Chemistry And Biochemistry
 The University of Oklahoma
 Center code:UOKNOR

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 5634: contig of 5634 bp in length
 * 5635 5734: gap of unknown length
 * 5735 14941: contig of 9207 bp in length
 * 14942 15041: gap of unknown length
 * 15042 26435: contig of 11394 bp in length
 * 26436 26536: gap of unknown length
 * 26536 58431: contig of 31896 bp in length
 * 58432 58531: gap of unknown length
 * 58532 109206: contig of 50675 bp in length.
 FEATURES
 source
 1..109206
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone_lib="ntb2-34p9"
 /clone_lib="Medicago truncatula BAC library H2"
 BASE COUNT 37237 a 17162 c 18248 g 36159 t 400 others
 ORIGIN
 Query Match 4.2%; Score 45.6; DB 2; Length 109206;
 Best Local Similarity 46.1%; Pred. No. 1.6;
 Matches 153; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
 QY 631 TTCTAATTCGAAGAATTTATGACTGCTTGATGGTCTAAGAAATCAGAAAGGAGACCGAA 690
 Db 23734 TTATGGTTCGGTTAAACCATTTGTTTACTAGTAACCAACCGATCCGACCAACAAAT 23675
 QY 691 AGCTTCTAAGTAATCAAGAAAAATCTTGACAAATAGCTTTTCCATAGAGAAAAACCGTGA 750
 Db 23674 ATTTTATCCAATGAATAAACCCACTTTTATTAATAAAAAATAAAAAATAACAAACTTT 23615
 QY 751 GGAAGGAGTCGATACCAATAATTCGATTAATAAAGGAGTCTTACACAGCCATATC 810
 Db 23614 AGGAGGAAGAAGTCCCTGTAAGTACAGTAAAGATTCTAATAAATCTCAATTTTAC 23555
 QY 811 AGTGAGCAACCTCCCTTTGATCTTTGATATTCATATGACAGAGATGACAAAGATAGTT 870
 Db 23554 TTATCTGCATCAGTCATAGATTGTTTGAAGTCTTGAACCAAAAAATAAATGATTAATA 23495
 QY 871 TTGAAAAATCTTAATAAGAAATGCAATTTGTCGAGCTAAAGCTTGCTCTCTTTATT 930
 Db 23494 TTACGATATATGTAATAGGTTTGATCTCATCTCAAAAAATAAATTAATATTATTTA 23435
 QY 931 TTCTTTTGTAGATGATTCGGTAGTAGGAAT 962
 Db 23434 TTGGATGTTTTAAATATTATTATTATTGGATT 23403
 RESULT 12
 AX083744
 LOCUS AX083744 1141 bp DNA linear PAT 28-FEB-2001
 DEFINITION Sequence 22 from Patent WO0111061.
 ACCESSION AX083744
 VERSION AX083744.1 GI:13185472
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 artificial sequences.

REFERENCE 1 (bases 1 to 1141)
 AUTHORS Kunst, L. and Clemens, S.
 TITLE Regulation of embryonic transcription in plants
 JOURNAL Patent: WO 011061-A 22 15-FEB-2001;
 UNIVERSITY OF BRITISH COLUMBIA (CA)
 FEATURES
 source
 1..1141
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 1..1141
 /note="consensus sequence of A.t., L.a., and B.n. FAEI
 promoters"
 BASE COUNT 123 a 32 c 42 g 112 t 832 others
 ORIGIN
 Query Match 4.1%; Score 45; DB 6; Length 1141;
 Best Local Similarity 11.0%; Pred. No. 3.6;
 Matches 114; Conservative 369; Mismatches 541; Indels 17; Gaps 2;
 QY 24 AAAGATATTTCTCTGATTGTAAGGGAATTCATTCTAACAGACCTTGGTCGAGCATCC 83
 Db 100 ARRWKDMKTAYBMTNTKNGKTGWRHRYWRWMBDVTVDHYYVTANNAATTTMCHMDKDD 159
 QY 84 ATTCTAATTTCAAGTCAGTCTTACAGATTCATCAAAATGCTCTTATGAATAAAGTCT 143
 Db 160 KTRWKKNNATGDDDTKYHNNNGCBTVWVRYKTDBDSBRKMYGHBWKNW 219
 QY 144 CTTGGAGGTGATGATTTGAAGGGAAGACGAATAATCTGATCCCAATCTTAAGGGGAA 203
 Db 220 STDVYYWVWDDCKRVRWRTGRMRNYWMBTAHRRYNNGTWBTAMAYRRWTMN 279
 QY 204 CAGTCAGTATCAATGAAGCGGTCTCAAGTTCCTGAGGAGAACTGATAAGAGCGCTC 263
 Db 280 NNNNAKAMCKRAKYMGWNRABVNSTCTTWKSKTKVTSWANNCRAGDANKDHKWKW 339
 QY 264 CATATGCTTTTCTCTTCATGCTTTTCAATAAGTCGCTTTCAACAGCTAAACCAAGC 323
 Db 340 SAAMGVYNNNNNNNTYKKAHBAWMDVHSAWKHANAHAHYSRKKWTBYRKTAVN 399
 QY 324 AGCTACAAATACGATCAGATTCGTTTGTAGGTTTGTATGCGTTGTTTCTTACTAGAA 383
 Db 400 NNGTMMKRWAWTWKMDMBGTYNNNNGRTYGYWTNKKMWTYKWKANNCKWRA 459
 QY 384 GAGCTTGACTTTTCTATGACTTTAGTGAAGAGTCTCCCATGTTTATGAANAATTCAG 443
 Db 460 WDHKTCTHNTTWKMKTY-----WNNCYWKSMTNGKSHREAAVYTWYMW 506
 QY 444 AAAATCAGGATGTAGAGCAAGCGACGTCCTCCCTAAAAGCTCATGATGATGATGAG 503
 Db 507 WRYAHANNNDWYHWRKACTWYKVCWKNNNTAAWYTKSWNTTSRYRWKTNNSRW 566
 QY 504 TTAAGAGAAATATCTTTATCGAGACTTCGCATTCAGAGGCAATGGGAAAGTAACGGT 563
 Db 567 RSDTSMGRANNYARABHYGYKWTWMBWHSHTWBHBRAGAAYHWMBMHYBAKCHMKAW 626
 QY 564 CTTGCAACAGAGACAGTATTTGATGGGAGCTACAAATCTGAAACGGGATAAATAGGAA 623
 Db 627 YKAKYAGAGSGSNNNNNNNNNNNNNNNNNNATC---ARDDYAAASRWYAMAKWYKBA 682
 QY 624 AATGTGTTTCTAATTCGAAGAAATTTAATGACTGCTTGGTCTAAGAAATCAGAAAGGA 683
 Db 683 ANNAYVTHANNWGCWNNTDTRRTMKNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 742
 QY 684 GACCGAAAGCTTCTAAGTAATCAAGAAAAATCTTCGACAATAGCTTTTCATAGAGAA 743
 Db 743 KAAKHWKANKWAMRGHADAABTTDKRNGAYTKYTTNNNTYRGVVTNTAARDGW 802
 QY 744 ACCGTCAGGAAGAGGTCGATACCAATAATTCGATTAATAAAGAGGAGTGTTCACACAG 803
 Db 803 ANNN 862
 QY 804 CCATATCAGTCGAGACCTTCCCTTCTGATCTTCTGATTAATGACAGAGATGACACAGA 863

Db	863	BAYTNNNNNNRMYGAYADDYAYYMSDTCDAWMKWDAATKNNNATYINRGTAWRNNNN	922
Db	864	GTAAGTTTTTGAAAAATTCATAAATAGAAATGCATTTGTGTGCGAGCTAAAGATCTGCTTCT	923
Db	923	NMTMTTKYBYBAAWNNNNNGKMCATHTWWVCATKTTKGCWNNCTTCTCKYKNNCTWT	982
Db	924	CTTTATTTCTCTTTGTAGAAATGATTCGGTGTAGTACGAATATGGGGTCGAGTAGTCACGCTT	983
Db	983	WMTTTTTAYAAATRWKTNATGSGMTRCNATGKNNNNYWTGKTRWTAYRATRWKAWK	1042
Db	984	GGAGTCATGGCTGTCGGAAGACAGGGAAGGATGTCGATCTTGTAGCAATCT	1043
Db	1043	VMATGSKWNTNSYRWKAYKTRAYKGWYNACAWRWGKATCYMTDNAAWTACATSWMATH	1102
Db	1044	TCAGATATACTTTGGGTCCA	1064
Db	1103	KYNWHCKNNNNNNNTMTR	1123

RESULT 13	AC024941	172246 bp	DNA	linear	PRI 11-JUL-2001
LOCUS	Human sapiens 12 BAC RP11-900F13	(Roswell Park Cancer Institute			
DEFINITION	Human BAC Library) complete sequence.				
ACCESSION	AC024941				
VERSION	AC024941.30	GI:14669928			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Carnivora; Catarrhini; Homnidae; Homo.				
	1 (bases 1 to 172246)				
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,				
	Alisbrooks,S.L., Amaraturunge,H.C., Are,J.R., Banks,T.J., Barbara,J.,				
	Benton,J., Blmage,M., Blankenburg,K., Bonnin,D., Bouck,J.,				
	Bowie,S., Brivega,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,				
	Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,				
	Cartier,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,				
	Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,				
	Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,				
	Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,				
	Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,				
	Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,				
	Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,				
	Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,				
	Garcia,A., Garner,T., Garza,N., Gill,R., Gottell,J.H., Guevara,W.,				
	Gonatane,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,				
	Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,				
	Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Honsi,F.,				
	Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,				
	Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,				
	Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,				
	Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,				
	Lewis,L.L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,				
	Louisleged,H., Lozaro,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,				
	Ma,J., Maheshwari,N., Mapua,P., Marondel,I., Martin,R.,				
	Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P.,				
	Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,				
	Miner,Z., Mitchell,I., Mohabbat,K., Montgomery,K.T., Morgan,N.,				
	Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,				
	Nguyen,S., Nguyen,N., Nguyen,N., Nickerson,E., Nockenkov,S.,				
	Ogih,M., Okuwono,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,				
	Peery,J., Perez,L., Rivers,M., Rojas,A., Rojubenik,I., Ruffe,M.,				
	Quiles,M., Ren,Y., Scher,S., Scott,G., Shen,H., Shim,C.,				
	Rulz,S., Savery,G., Scher,S., Scott,G., Shen,H., Shim,C.,				
	Shoostari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A.,				
	Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,				
	Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,				
	Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,				
	Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,				
	Washington,C., Watlington,S., Williams,G., Williamson,A.,				
	Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,				
	Zorrilla,S., Kucherlapati,R. and Gibbs,R.				

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
 Unpublished
 2 (bases 1 to 172246)
 Worley,K.C.
 Direct Submission
 Submitted (03-MAR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 172246)
 Worley,K.C.
 Direct Submission
 Submitted (11-JUL-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2001 this sequence version replaced gi:13492990.
INFORMATION: <http://hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:
 STSS are identified using ePCR (Genome Res. 7:541-550), searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.
 Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.
 Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

COMMENT

QUALSTAT-REPORT-----

```

----- Summary Statistics -----
Contig length:
Phrap values in estimate:
Average error rate (BCM-Phrap estimate):
Fraction of Phrap values less than 40 :
Number of consensus changing edits:
Number of N's in consensus :

```

Position	Original context	Consensus changing edits	Edited+Context
67598	tctcagacag(n)aattgacaata		tctcagacag(c)aattgacaata
67599	ctctttgaa(n)tattattatg		ctctttgaa(c)tattattatg
99512	gcatctggtg(n)nacagacccc		gcatctggtg(c)nacagacccc
103635	gcatctggtg(n)nacagacccc		gcatctggtg(c)nacagacccc
103636	gcatctggtg(n)nacagacccc		gcatctggtg(c)nacagacccc
103660	aaacatctgga(n)ccagatgttg		aaacatctgga(a)ccagatgttg
106704	tatatctacct(a)ataaaatnt		tatatctacct(t)ataaaatnt
106705	gattctacct(a)ataaaatnt		gattctacct(t)ataaaatnt

* 4054 4783: contig of 730 bp in length
* 4784 gap of unknown length
* 4884 5571: contig of 688 bp in length
* 5572 5671: gap of unknown length
* 5672 6376: contig of 705 bp in length
* 6377 6476: gap of unknown length
* 6477 7207: contig of 731 bp in length
* 7208 7307: gap of unknown length
* 7309 7613: contig of 306 bp in length
* 7614 7713: gap of unknown length
* 7715 8902: contig of 1189 bp in length
* 8903 9757: contig of 755 bp in length
* 9758 9857: gap of unknown length
* 9858 10343: contig of 486 bp in length
* 10344 11201: contig of 758 bp in length
* 11202 12335: contig of 1034 bp in length
* 12336 12435: gap of unknown length
* 12436 13294: contig of 859 bp in length
* 13295 13394: gap of unknown length
* 13395 14096: contig of 702 bp in length
* 14097 15192: contig of 1056 bp in length
* 15193 16686: gap of unknown length
* 16687 17422: contig of 1314 bp in length
* 17423 17522: gap of unknown length
* 17523 18416: contig of 636 bp in length
* 18417 18516: gap of unknown length
* 18517 19058: contig of 894 bp in length
* 19059 20107: contig of 542 bp in length
* 20108 20207: gap of unknown length
* 20209 21392: contig of 949 bp in length
* 21393 21492: contig of 1185 bp in length
* 21493 22510: gap of unknown length
* 22511 23648: contig of 1018 bp in length
* 23649 23748: gap of unknown length
* 23749 24975: contig of 1038 bp in length
* 24976 25075: contig of 1227 bp in length
* 25076 26075: gap of unknown length
* 26076 26175: contig of 1000 bp in length
* 26176 27416: gap of unknown length
* 27417 27516: contig of 1241 bp in length
* 27517 28529: contig of 1013 bp in length
* 28530 28629: gap of unknown length
* 28630 29583: contig of 954 bp in length
* 29584 30741: gap of unknown length
* 30742 30841: contig of 1058 bp in length
* 30842 33988: gap of unknown length
* 33989 34089: contig of 1148 bp in length
* 34090 34154: gap of unknown length
* 34155 34254: contig of 1065 bp in length
* 34255 34354: gap of unknown length
* 34355 34457: contig of 1203 bp in length
* 34458 34557: gap of unknown length
* 34559 35986: contig of 1429 bp in length
* 35987 36086: gap of unknown length
* 36087 37925: contig of 1839 bp in length
* 37926 38025: gap of unknown length
* 38026 39294: contig of 1269 bp in length
* 39295 39394: gap of unknown length
* 39395 40532: contig of 1138 bp in length
* 40533 40632: gap of unknown length
* 40633 42129: contig of 1497 bp in length
* 42130 42229: gap of unknown length
* 42230 43855: contig of 1626 bp in length
* 43856 43955: gap of unknown length
* 43956 45742: contig of 1787 bp in length

45743
45843 gap of unknown length
45844 contig of 957 bp in length
46899 gap of unknown length
46900 contig of 1410 bp in length
46901 gap of unknown length
48310 48409: contig of 1359 bp in length
48410 49768: gap of unknown length
49769 49868: gap of unknown length
51572 contig of 1804 bp in length
51573 51772: gap of unknown length
51773 53070: contig of 1298 bp in length
53071 53170: gap of unknown length
53171 55690: contig of 2720 bp in length
55691 55990: gap of unknown length
55991 58340: contig of 2550 bp in length
58341 58440: gap of unknown length
58441 60742: contig of 2102 bp in length
60743 60842: gap of unknown length
60843 62552: contig of 1710 bp in length
62553 65163: gap of unknown length
65164 65263: gap of unknown length
65264 66884: contig of 1721 bp in length
66885 67084: gap of unknown length
67085 68550: contig of 1466 bp in length
68551 68650: gap of unknown length
68651 70725: contig of 2075 bp in length
70726 70825: gap of unknown length
70826 72615: contig of 1790 bp in length
72616 72715: gap of unknown length
72716 74764: contig of 2049 bp in length
74765 74864: gap of unknown length
74865 77074: contig of 2210 bp in length
77075 77174: gap of unknown length
77175 79340: contig of 2166 bp in length
79341 79440: gap of unknown length
79441 80990: contig of 1550 bp in length
80991 81090: gap of unknown length
81091 82975: contig of 1885 bp in length
82976 83075: gap of unknown length
83077 84926: contig of 1851 bp in length
84927 85026: gap of unknown length
85027 87534: contig of 2528 bp in length
87535 87634: gap of unknown length
87635 90562: contig of 2908 bp in length
90563 90662: gap of unknown length
90663 92287: contig of 1624 bp in length
92288 92386: gap of unknown length
92387 95351: contig of 2965 bp in length
95352 95451: gap of unknown length
95452 97620: contig of 2169 bp in length
97621 97720: gap of unknown length
97721 100394: contig of 2674 bp in length
100395 100494: gap of unknown length
100495 102562: contig of 2068 bp in length
102563 102662: gap of unknown length
102663 105607: contig of 2945 bp in length
105608 105708: gap of unknown length
105709 108635: contig of 2828 bp in length
108636 111168: gap of unknown length
111169 111268: contig of 2533 bp in length
111269 114040: gap of unknown length
114041 114140: contig of 2772 bp in length
114141 117074: contig of 2933 bp in length
117075 117174: gap of unknown length
117175 119836: contig of 2663 bp in length
119837 119936: gap of unknown length
119937 123089: contig of 3153 bp in length
123090 123189: gap of unknown length
123190 125696: contig of 2507 bp in length
125697 125797: gap of unknown length
125798 128622: contig of 2826 bp in length
128623 128722: gap of unknown length

750 AGGAAAGGAGTCGATACCA"AAATTCCGATTAAAAA 786

QY	750	AACATTCGCTCGAATTGGTTTCCCAAGCCTCATGCATG
Db	222544	ATAAAAAAATTTTTAAATTAATAACCAAAAAAAAAATAAAAA
	222508	

RESULT	18
ACQ008008	
LOCUS	Homo sapiens Xp22 PAC RPcG16-102 linear PRI 16-MAY-2002
DEFINITION	(Roswell Park Cancer Institute)
DESCRIPTION	Human PAC Library) complete sequence.
ACCESSION	ACQ008008
VERSION	ACQ008008.2 GI:56556683
KEYWORDS	HTC.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens.

ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 167257)
AUTHORS	Muzny, D., Aronson, A.D., Bouck, J., Brundage, E., Bunac, C., Chen, Z., Di, W., Ding, Y., Dugan, S., Durbin, J., Forcum, J., Garcia, C., Gorrell, J.H., Gorrell, L.L., Hernandez, J., Jackson, L., Kondeljewski, N., Leal, B., Lichter, O., Liu, W., Logan, O., Lu, J., Martinez, C., Oswal, G., Pampell, L.R., Parish, B.J., Perez, L., Rashid, N.D., Rives, C., Scherer, S.E., Shen, H., Simon, M., Vo, Q., Williamson, A., Worley, K.C., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.A.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 167257)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 167257)

AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 167257)

AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 167257)

Worley, K.C.
Direct Submission
Submitted (09-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6 (Phases 1 to 167257)

COMMENTS:
 Worley, K. C.
 Direct Submission
 Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 30, 1999 this sequence version replaced gi:5441911.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
SVSSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished), for Human and Mouse sequences. Genes and Regions 32389-34021 are identified by BLAST (Nuc. Acids Res. 32389-34021) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not continuously mapped are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

	Summary Statistics
Contig length:	167257
Phrap values in estimate:	166248
Average error rate (BCM-Phrap estimate):	0.000259222
Fraction of Phrap values less than 40:	0.0376546
Number of consensus changing edits:	28
Number of N's in consensus:	0

Position	Consensus changing edits	Original-Context	Edited-Context
11124	ttgtgtgac(n)atnnnncatg	ttgtgtgac(n)gtatgtcacatg	ttgtgtgac(n)gtatgtcacatg
11128	gatgacnatt(n)ncatgaaat	gatgacatt(g)acatgaagaat	gatgacatt(g)acatgaagaat
11129	atgacnatt(n)ncatgaatc	atgacatt(g)acatgaagaatc	atgacatt(g)acatgaagaatc
11130	tgacnatt(n)catgaatacat	tgacattatg(n)atgaatacatc	tgacattatg(n)atgaatacatc
11660	taagtgtctg(n)agatgcagtt	aaaggtgtcag(n)agatgcagtt	aaaggtgtcag(n)agatgcagtt
32920	aaaggtgtcag(n)agatgcagtt	acttgcattct(t)tcacatccag	acttgcattct(t)tcacatccag
32998	tgacattctt(n)ccagttatatt	tgacattctt(n)ccagttatatt	tgacattctt(n)ccagttatatt
33125	ttttctttcc(n)tttttttttt	ttttctttcc(n)tttttttttt	ttttctttcc(n)tttttttttt
64695	ttttctttcc(n)tttttttttt	ttttctttcc(n)tttttttttt	ttttctttcc(n)tttttttttt
70901	tctacatca(n)ataataaag	tctacatca(n)ataataaag	tctacatca(n)ataataaag
72799	cacgacatta(n)ttgaataaga	cacgacatta(n)ttgaataaga	cacgacatta(n)ttgaataaga
119977	aattaccaaa(n)gtgngggggn	aattaccaaa(n)gtgngggggn	aattaccaaa(n)gtgngggggn
119982	ccaaagtctg(n)gggngcagg	ccaaagtctg(n)gggngcagg	ccaaagtctg(n)gggngcagg
119987	ngtngggg(n)nccagatttga	tgtgngggggt(t)ccagatttga	tgtgngggggt(t)ccagatttga
120157	caaacatt(n)ggggtttaaa	caaacatttctg(n)ggggtttaaa	caaacatttctg(n)ggggtttaaa
120195	tgtgtgtctg(n)agatgagcta	tgtgtgtctg(n)agatgagcta	tgtgtgtctg(n)agatgagcta
121800	aattaccan(n)aaagttagac	aattaccan(n)aaagttagac	aattaccan(n)aaagttagac
121881	attaccacn(n)aaagttagac	attaccacn(n)aaagttagac	attaccacn(n)aaagttagac
142786	agaaatctgga(n)aaaataatt	agaaatctgga(n)aaaataatt	agaaatctgga(n)aaaataatt
141954	ataagttagt(n)tgatttcaatt	ataagttagt(n)tgatttcaatt	ataagttagt(n)tgatttcaatt
150222	ttaattctgat(n)tgctctctct	ttaattctgat(t)tgctctctct	ttaattctgat(t)tgctctctct
150300	ctggagata(n)ctaaataacag	ctggagata(n)ctaaataacag	ctggagata(n)ctaaataacag
150332	aaataccttt(n)cataagacat	aaataccttt(n)cataagacat	aaataccttt(n)cataagacat
150354	ccacacaggt(n)gccattgca	ccacacaggt(n)gccattgca	ccacacaggt(n)gccattgca
150594	atggagctgt(n)acactgtgc	atggagctgt(n)acactgtgc	atggagctgt(n)acactgtgc
160398	tgatgtgtct(n)ttmnagagat	tgatgtgtct(n)ttmnagagat	tgatgtgtct(n)ttmnagagat
160400	tttggtctct(n)naggagatta	tttggtctct(n)naggagatta	tttggtctct(n)naggagatta
160401	tttggtctct(n)naggagattta	tttggtctct(n)naggagattta	tttggtctct(n)naggagattta

```

----- Distribution of Quality < 40 Bases -----
#
1000|      *
900|      *
800|      *
700|      *

```


This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPI1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J. J., and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-569B21, 2000 bp overlap; the clone sequenced to the right is RP11-727O11. Actual start of this clone is at base position 1 of RP11-510H11; actual end is at base position 198230 of RP11-510H11.

The sequence of AC026578 has been incorporated into AC092662.

FEATURES

```

source
    1. .198230
       Location/Qualifiers
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /chromosome="2"
       /map="2"
       /clone="RP11-510H11"
       /clone_lib="RPI1-11"
       repeat_region
         151..265
         /rpt_family="MIR"
       repeat_region
         288..328
         /rpt_family="L2"
       repeat_region
         364..3683
         /rpt_family="AT_rich"
       repeat_region
         3782..4214
         /rpt_family="ERV1"
       repeat_region
         4246..4282
         /rpt_family="AT_rich"
       repeat_region
         5137..5508
         /rpt_family="L1"
       repeat_region
         5509..5804
         /rpt_family="Alu"
       misc_feature
         5786..5804
         /note="similar to Homo sapiens EST BE874285
         (NID:g10323061)"
       repeat_region
         5805..6602
         /rpt_family="L1"
       repeat_region
         7076..7108
         /rpt_family="AT_rich"
       misc_feature
         7355..7549
         /note="similar to Homo sapiens EST BE907740
         (NID:g10401605)"
       misc_feature
         7409..7549
         /note="similar to Homo sapiens EST BE874285
         (NID:g10323061)"
       misc_feature
         7683..7991
         /note="similar to Homo sapiens EST BE907740
         (NID:g10401605)"
       misc_feature
         7683..7975
         /note="similar to Homo sapiens EST AW993282
         (NID:g8253434)"

```

```

misc_feature
    7683..7975
    /note="similar to Homo sapiens EST BG253627
    (NID:g12763443)"
misc_feature
    7683..7975
    /note="similar to Homo sapiens EST BG499781
    (NID:g13461238)"
misc_feature
    7683..7843
    /note="similar to Homo sapiens EST BE874285
    (NID:g10323061)"
repeat_region
    8827..8861
    /rpt_family="(TTTA)n"
repeat_region
    8846..9142
    /rpt_family="Alu"
repeat_region
    9162..9243
    /rpt_family="L2"
repeat_region
    9347..9395
    /rpt_family="(TTTA)n"
repeat_region
    9383..10565
    /rpt_family="L1"
repeat_region
    11775..12096
    /rpt_family="L2"
repeat_region
    12152..12363
    /rpt_family="MaLR"
repeat_region
    12421..12514
    /rpt_family="MaLR"
repeat_region
    12528..13282
    /rpt_family="ERV1"
repeat_region
    13307..13351
    /rpt_family="L2"
repeat_region
    13477..13519
    /rpt_family="(CA)n"
repeat_region
    15066..16328
    /rpt_family="L1"
repeat_region
    16329..16802
    /rpt_family="L1"
repeat_region
    16587..16618
    /rpt_family="AT_rich"
repeat_region
    16603..16761
    /rpt_family="L1"
repeat_region
    16770..16817
    /rpt_family="ERV1"
repeat_region
    16797..16899
    /rpt_family="(TAGA)n"
repeat_region
    16936..17007
    /rpt_family="MER2_type"
repeat_region
    17008..17085
    /rpt_family="MER2_type"
repeat_region
    17092..17417
    /rpt_family="L1"
repeat_region
    17423..17913
    /rpt_family="L1"
repeat_region
    17912..17970
    /rpt_family="L1"
repeat_region
    17971..18267
    /rpt_family="Alu"
repeat_region
    18268..19599
    /rpt_family="L1"
repeat_region
    19560..19697
    /rpt_family="AT_rich"
repeat_region
    19860..19921
    /rpt_family="MIR"
repeat_region
    20016..20038
    /rpt_family="AT_rich"
repeat_region
    22060..22080
    /rpt_family="AT_rich"
repeat_region
    23469..23526
    /rpt_family="ERV1"
repeat_region
    24064..24427
    /rpt_family="MaLR"
repeat_region
    24567..25448

```

Query Match 4.1% Score 44; DB 9; Length 198230;
Best Local Similarity 46.4%; Pred. No. 3.5;

Db	1230	AAAATCCTAAAAAATAAATTAATATATAACAAATAAATTTTTAAAAACCAAAAAACAA	1171
QY	801	CAGCCATATCAGTGGCAGCAGCTTCCTCTTTGATCTTTTGAATATGACAGAATGACAA	860
Db	1170	AACCAGGAAATTCAAATACCTAAATATTAAGCGTTAAAAAATAAATATACAA	1111
QY	861	AGAGTAAGTTTTGAAAAATCTAAATAGAAATGCATTTCTCTCGAGCTAAAGCTTGC	920
Db	1110	AAAAACCTCAATAAAAAATCTTAAAAACCAACAACATATTTAAAAAATAACCAATATTACA	1051
QY	921	TCTCTTAATTTTCCCTTTTGTAGAT	945
Db	1050	TAATCAAAATTTTATTTTCAAAAAAT	1026

[illegible]

```

Query Match          4.0%; Score 43.2; DB 2; Length 88966;
Best Local Similarity 52.2%; Pred. No. 5.9;
Matches 96; Conservative 88; Indels 0; Gaps 0
BASE COUNT 27101 a 15563 c 16501 g 29198 t 603 others
ORIGIN
/clone="x18399"

```

QY	618	AGGAAATGCTTCTAATTCGAGAAATTTAATGACTGCTTTAGTGCTTAAGAAATCAG	677
Db	63048	AGCATATGAGGATTAGAAATTAACAGCTGAATAGAAATTTACAATTTTAAAGAAATCAA	62989
QY	678	AAAGGAGACCGAAGCTTCTAAGTATCAAGAAAAAATCTTGGACAATAGCTTTTCCATA	737
Db	62988	CCAGCTTAAGAGTASCTTTTGGAAAGATCAACAATACTAACAGACATTTATCTATATG	62929
QY	738	GAGAAACCGTGAGAAAGAGCTGATACCATTAATCCGATTTAAAAAGSGAGTGCTTA	797
Db	62928	AGAAAAAAAAGAGAAAGATTTCAATGATTAATACATATGAAAGGGGACATTA	62869
QY	798	CAAC 801	
Db	62868	CAAC 62865	

RESULT	23	AC093810	146671 bp	DNA	linear	PRI 01-MAR-2002
LOCUS	AC093810/c					
DEFINITION		Homo sapiens BAC clone RP11-354024 from 4, complete sequence.				
ACCESSION	AC093810	AC023636				
VERSION	AC093810.4	GI:18042417				
KEYWORDS	HTG.					
SOURCE	Homo sapiens.					
ORGANISM	Homo sapiens					
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 146671)				
TITLE		Sulston, J.E. and Waterston, R.				
JOURNAL		Toward a complete human genome sequence				
MEDLINE		Genome Res. 8 (11), 1097-1108 (1998)				
PUBMED		99063792				
REFERENCE		9847074				
AUTHORS		2 (bases 1 to 146671)				
TITLE		Kalicki, J. and Haglund, K.				
JOURNAL		The sequence of Homo sapiens BAC clone RP11-354024				
REFERENCE		Unpublished (2001)				
AUTHORS		3 (bases 1 to 146671)				
TITLE		Waterston, R.H.				
JOURNAL		Direct Submission				
REFERENCE		Submitted (10-SEP-2001) Genome Sequencing Center, Washington				
AUTHORS		University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
TITLE		4 (bases 1 to 146671)				
JOURNAL		Waterston, R.H.				
REFERENCE		Direct Submission				
AUTHORS		Submitted (03-JAN-2002) Genome Sequencing Center, Washington				
TITLE		University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA				
JOURNAL		5 (bases 1 to 146671)				
REFERENCE		Waterston, R.				
AUTHORS		Direct Submission				
TITLE		Submitted (01-MAR-2002) Department of Genetics, Washington				
JOURNAL		University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
COMMENT		On Jan 3, 2002 this sequence version replaced gi:16077045.				
		----- Genome Center				
		Center: Washington University Genome Sequencing Center				
		Center code: WUGSC				
		Web site: http://genome.wustl.edu/gsc				
		Contact: sapiens@watson.wustl.edu				
		----- Summary Statistics				
		Center project name: H_NH0354024				
		Drafting Center: WITB				

	FEATURES	SOURCE
1.	Number of employees	U.S. Bureau of Economic Analysis
2.	Value added	"
3.	Capital expenditures	"
4.	Research and development	"
5.	Advertising	"
6.	Depreciation	"
7.	Dividends	"
8.	Earnings before interest and taxes	"
9.	Interest expense	"
10.	Income tax expense	"
11.	Net income	"
12.	Operating expenses	"
13.	Prepaid expenses	"
14.	Provisions for doubtful accounts	"
15.	Retained earnings	"
16.	Total assets	"
17.	Total liabilities	"
18.	Total equity	"
19.	Working capital	"
20.	Current ratio	"
21.	Debt-to-equity ratio	"
22.	Return on assets	"
23.	Return on equity	"
24.	Profit margin	"
25.	Turnover of assets	"
26.	Turnover of equity	"
27.	Turnover of sales	"
28.	Days receivable	"
29.	Days payable	"
30.	Inventory turnover	"
31.	Fixed asset turnover	"
32.	Capital turnover	"
33.	Equity turnover	"
34.	Debt turnover	"
35.	Asset turnover	"
36.	Equity turnover	"
37.	Debt turnover	"
38.	Asset turnover	"
39.	Equity turnover	"
40.	Debt turnover	"
41.	Asset turnover	"
42.	Equity turnover	"
43.	Debt turnover	"
44.	Asset turnover	"
45.	Equity turnover	"
46.	Debt turnover	"
47.	Asset turnover	"
48.	Equity turnover	"
49.	Debt turnover	"
50.	Asset turnover	"
51.	Equity turnover	"
52.	Debt turnover	"
53.	Asset turnover	"
54.	Equity turnover	"
55.	Debt turnover	"
56.	Asset turnover	"
57.	Equity turnover	"
58.	Debt turnover	"
59.	Asset turnover	"
60.	Equity turnover	"
61.	Debt turnover	"
62.	Asset turnover	"
63.	Equity turnover	"
64.	Debt turnover	"
65.	Asset turnover	"
66.	Equity turnover	"
67.	Debt turnover	"
68.	Asset turnover	"
69.	Equity turnover	"
70.	Debt turnover	"
71.	Asset turnover	"
72.	Equity turnover	"
73.	Debt turnover	"
74.	Asset turnover	"
75.	Equity turnover	"
76.	Debt turnover	"
77.	Asset turnover	"
78.	Equity turnover	"
79.	Debt turnover	"
80.	Asset turnover	"
81.	Equity turnover	"
82.	Debt turnover	"
83.	Asset turnover	"
84.	Equity turnover	"
85.	Debt turnover	"
86.	Asset turnover	"
87.	Equity turnover	"
88.	Debt turnover	"
89.	Asset turnover	"
90.	Equity turnover	"
91.	Debt turnover	"
92.	Asset turnover	"
93.	Equity turnover	"
94.	Debt turnover	"
95.	Asset turnover	"
96.	Equity turnover	"
97.	Debt turnover	"
98.	Asset turnover	"
99.	Equity turnover	"
100.	Debt turnover	"
101.	Asset turnover	"
102.	Equity turnover	"
103.	Debt turnover	"
104.	Asset turnover	"
105.	Equity turnover	"
106.	Debt turnover	"
107.	Asset turnover	"
108.	Equity turnover	"
109.	Debt turnover	"
110.	Asset turnover	"
111.	Equity turnover	"
112.	Debt turnover	"
113.	Asset turnover	"
114.	Equity turnover	"
115.	Debt turnover	"
116.	Asset turnover	"
117.	Equity turnover	"
118.	Debt turnover	"
119.	Asset turnover	"
120.	Equity turnover	"
121.	Debt turnover	"
122.	Asset turnover	"
123.	Equity turnover	"
124.	Debt turnover	"
125.	Asset turnover	"
126.	Equity turnover	"
127.	Debt turnover	"
128.	Asset turnover	"
129.	Equity turnover	"
130.	Debt turnover	"
131.	Asset turnover	"
132.	Equity turnover	"
133.	Debt turnover	"
134.	Asset turnover	"
135.	Equity turnover	"
136.	Debt turnover	"
137.	Asset turnover	"
138.	Equity turnover	"
139.	Debt turnover	"
140.	Asset turnover	"
141.	Equity turnover	"
142.	Debt turnover	"
143.	Asset turnover	"
144.	Equity turnover	"
145.	Debt turnover	"
146.	Asset turnover	"
147.	Equity turnover	"
148.	Debt turnover	"
149.	Asset turnover	"
150.	Equity turnover	"
151.	Debt turnover	"
152.	Asset turnover	"
153.	Equity turnover	"
154.	Debt turnover	"
155.	Asset turnover	"
156.	Equity turnover	"

BASE COUNT
ORIGIN

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPII-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is AC087607; the clone sequenced to the right is RP11-484F3, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-354024; actual end is at base position 34976 of RP11-484F3.

There is an ambiguous at base position 5636.

The sequence of AC023636 has been incorporated into AC093810.

FEATURES

Source

Location/Qualifiers

1..146671

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-354024"

/clone_lib="RPII-11"

1..239

/rpt_family="ERV1"

341..503

/rpt_family="ERV1"

515..667

/rpt_family="ERV1"

681..1056

/rpt_family="ERV1"

1058..1534

/rpt_family="L1"

1644..1728

/rpt_family="MIR"

1746..2036

/note="match to EST T41120 (NID:9648688) ya31a05.s2"

1943..2237

/note="match to EST T40287 (NID:9647923) ya31e05.r2"

2872..2916

/rpt_family="AT_rich"

3714..4105

/rpt_family="L1"

4121..4623

/rpt_family="L1"

4199..4220

/rpt_family="AT_rich"

4638..4865

/rpt_family="L1"

4891..6849

/rpt_family="L1"

5591..5702

/rpt_family="(TA)n"

6532..6553

/rpt_family="AT_rich"

6748..6776

/rpt_family="(TAAA)n"

repeat_region 7416..7668
/rpt_family="L2"
repeat_region 8370..8483
/rpt_family="ACHobo"
repeat_region 8562..8954
/rpt_family="L1"
repeat_region 8968..9041
/rpt_family="MIR"
repeat_region 15112..15282
/rpt_family="MIR"
repeat_region 15296..15416
/rpt_family="L1"
repeat_region 15469..15706
/rpt_family="L1"
repeat_region 15759..15889
/rpt_family="Alu"
repeat_region 16136..16229
/rpt_family="L1"
repeat_region 17713..18027
/rpt_family="L1"
repeat_region 18028..18616
/rpt_family="L1"
repeat_region 18782..18840
/rpt_family="AT_rich"
repeat_region 19690..19949
/rpt_family="Alu"
repeat_region 20271..20872
/rpt_family="L2"
repeat_region 20903..21256
/rpt_family="L2"
repeat_region 21531..22139
/rpt_family="ERV1"
repeat_region 22360..22425
/rpt_family="L2"
repeat_region 22477..22646
/rpt_family="CR1"
repeat_region 22953..22993
/rpt_family="(TC)n"
repeat_region 24268..24346
/rpt_family="MIR"
repeat_region 24359..24406
/rpt_family="AT_rich"
repeat_region 26984..27033
/rpt_family="AT_rich"
repeat_region 27186..27799
/rpt_family="L2"
repeat_region 28814..28850
/rpt_family="AT_rich"
repeat_region 28906..29115
/rpt_family="L1"
repeat_region 29116..29149
/rpt_family="MaLR"
repeat_region 29150..29507
/rpt_family="L1"
repeat_region 29505..29527
/rpt_family="AT_rich"
repeat_region 30463..30505
/rpt_family="AT_rich"
repeat_region 30555..31651
/rpt_family="L1"
repeat_region 31663..32733
/rpt_family="L1"
repeat_region 32109..32133
/rpt_family="AT_rich"
repeat_region 32734..33108
/rpt_family="MaLR"
repeat_region 33109..33153
/rpt_family="L1"
repeat_region 33622..33908
/rpt_family="Alu"
repeat_region 33907..33956
/rpt_family="(TAA)n"
repeat_region 33937..33971

	Query Match	4.0%	Score 43.2	DB 9	Length 146671
	Best Local Similarity	52.2%	Pred. No. 5.6		
	Matches	96	Conservative	0	Mismatches 88; Indels 0; Gaps 0
618	AGGAAAAATGTTTCCTAAATTCGAAGAATTTTAAATGACTGCTGTGATGGCTCTAAGAAATCAG	677			
42708	AGAAATAATGAAGATTAGCAATTAAGAAGCAGTGAATACAAATTTACAAATTTTAAAGAATCAA	42649			
678	AAGGAGACCCGAAAGCTTCCTAAGCTAATCAAGAAAAAATCTGTGCAACAATAGCTTTTCCATA	737			
42648	CCAAGCTAAGGTAGGTTTTTTTGAAGAATCAACAAAACCTAACAGACATTTATCTATATG	42589			
738	GAGAAACCGTGGAGGAAGGTCGATACCAATAATTCGGATTAAAAAAGGGAGTGCTTA	797			
42588	AAGAAAAAAGAGAAAAAGATTCAANTGGATAAAATACATATGAAGAAAGGGGACATTA	42529			
798	CAAC 801				
42528	CAAC 42525				
RESULT 24					
AC107307/c		158066 bp	DNA	linear	PRI 31-JUL-2002
LOCUS	Homo sapiens 3 BAC RP11-657G2 (Roswell Park Cancer Institute Human				
DEFINITION	BAC Library) complete sequence.				
ACCESSION	AC107307				
VERSION	AC107307.3	GI:19774327			
KEYWORDS	HTG.				
SOURCE	Human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	1 (bases 1 to 158066)				
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,				
	Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,				
	Barbarta,J., Benton,J., Binage,K., Blankenburg,K., Bonin,D.,				
	Bouck,J., Bowie,S., Briveva,M., Brown,E., Brown,M., Bryant,N.P.,				
	Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,				
	Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,				
	Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,				
	Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,				
	Delavelle,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,				
	DeLanay,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.J.,				
	Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.H.,				
	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,				
	Falls,T., Ferraguto,D., Flagg,N., Ford,J.J., Foster,P., Frantz,P.,				
	Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,				
	Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,				
	Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,				
	Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,				
	Hollins,B., Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,				
	Jackson,L.B., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,				
	Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,				
	Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,				
	Levis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,				
	Loulsheed,H., Lozaro,R.J., Lu,X., Lucier,A., Luciet,R., Luna,R.,				
	Ma,J., Maheshwari,M., Mapu,P., Martin,K., Martindale,A.,				
	Martinez,E., Massey,E., Mathiney,E., McLeod,M.P., Meador,M.,				
	Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,				
	Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,				
	Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,				
	Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,				
	Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,				
	Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,				
	Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,				
	Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E.,				
	Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,				
	Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,				
	Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,				
	Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,				
	Warren,R., Washington,C., Watlington,S., Williams,G.,				

Williamson, A., Wleczyk, R., Woodede, S., Worley, K., Wu, C., Wu, Y.,
Wu, F., Zhou, J., Zorrilla, S., Gaylor, S.L., Weinstock, G. and
Gibbs, R.
Direct Submission
Unpublished
2. (bases 1 to 158066)
Worley, K.C.
Direct Submission
Submitted (18-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3. (bases 1 to 158066)
Worley, K.C.
Direct Submission
Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4. (bases 1 to 158066)
Worley, K.C.
Direct Submission
Submitted (28-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5. (bases 1 to 158066)
Worley, K.C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
6. (bases 1 to 158066)
Worley, K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Mar 28, 2002 this sequence version replaced gi:18449849.
Information: <http://www.hgsc.bcm.tmc.edu/> or email
cb-hel@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

ANNOTATION OF FEATURES:
STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity are identified by BLAST (unpublished.) for Human and Mouse sequences.

Genes and region of sequence similarity (expect < 1e-34) to the (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

FEATURES Location/Qualifiers

source 1. .158066

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="3"

/clone="RP11-657G2"

misc_feature 1. .2000

/note="overlaps bases 17989. .19988 of clone AC117506"

/function="clone overlap"

repeat_region 136. .686

repeat_region 717. .923

repeat_region complement(958. .1128)

repeat_region 1252. .1799

repeat_region 3330. .3367

repeat_region 3519. .3940

repeat_region 4242. .4287

repeat_region 4421. .4479

repeat_region 4480. .4513

repeat_region complement(4514. .4790)

repeat_region 4895. .4924

repeat_region 4951. .4985

repeat_region 5020. .5041

repeat_region 5079. .5504

repeat_region 5508. .6426

repeat_region 6374. .6818

repeat_region 6813. .8139

repeat_region 8140. .8445

repeat_region 8446. .8679

repeat_region 8686. .8997

repeat_region 9528. .9955

repeat_region 10136. .10175

repeat_region complement(10229. .10588)

repeat_region 10229. .10588

repeat_region complement(10700. .10996)

repeat_region 11042. .11079

repeat_region 14756. .14800

repeat_region 16709. .17324

repeat_region 19258. .19280

repeat_region 19658. .19680

repeat_region complement(19837. .20082)

repeat_region 20605. .22071

repeat_region complement(23423. .23578)

repeat_region 25577. .25684

repeat_region complement(27105. .27211)

Query Match 3.9% Score 42.8; DB 9; Length 158066;

Best Local Similarity 45.3%; Pred. No. 6.8;

Matches 155; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 607 ACACGGATAATAGAAAATGTGTTCTTAATTCGAGAAATTTAATGACTGCTTGTGATGTC 666

DB 79281 AATTCGACATATGAACATTAAGGAATCCAAATATGAATAGGCTGTGGCAG 79222

QY 667 TAAGAAATCAGAAAGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAATCTTGGACAATA 726

DB 79221 ACTTTCAGATCTAATGCCATTTATTTTCATTTGTCAGTCAGATATAAAAGTAACATTGTA 79162

QY 727 GCTTTTCCATAGAGAAAACCGTGAGGAAAGAGTCGATACCATATAATCCGATTAAAAAA 786

DB 79161 CATTAACTTAAATATAACAGAAAGAAAGAGAAAGAAAGAAAGAAAGAAAGAA 79102

QY 787 GGGAGTGTCTTACAACAGCCATATCATGTCGAGCAACCTTCCCTTTGATTCTTTGATAATGA 846

DB 79101 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 79042

QY 847 CAGAAGATGAACAAGAGTAAAGTTTGTAAATTTCTAAATAGAAATGCAATTTGTCTCG 906

DB 79041 AGTATACCTTTATAGATAATTTTATAGAAAATTAATTAACATATAAATATTAGATTG 78982

QY 907 ACCTAAAGCTTGTCTTCTTTATTTTCCCTTTTGTAGATGAT 948

DB 78981 ACCTAATACATCCCACTATATGATTATTTTACTTTAAAGCAT 78940

RESULT 25

AC025807

LOCUS

DEFINITION Homo sapiens clone RP11-657G2, WORKING DRAFT SEQUENCE, 15 unordered pieces.

AC025807

VERSION AC025807.2 GI:7342200

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 159716)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-657G2

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 159716)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Menes, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talanas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186158)
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P., Brottier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F., Levy, M., Eckenberg, R., Bruls, T., Gherardinis, V., Cruaud, C., Gvayap, G., Saurin, W. and Weissenbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 186158)
Genoscope.
Direct Submission
Submitted (26-APR-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
On Feb 28, 2001 this sequence version replaced gi:11611112.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-66E7
Downstream BAC (overlapping the SP6 end) : R-423117 (AC=AL513142)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.54x in Q20 bases; sum-of-contigs

Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 :
30 - 39 :
40 - 49 :
50 - 59 :
60 - 69 :
70 - 79 :
80 - 89 :
90 - 99 :

50
1758
4042
8680
23674
59795
88159

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
Source
1. 186158
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-60347"
/clone_lib="RPC1-11"
146288..146422
/note="matching EMBL:238793
RHdb:RH53848
dbSTS:STS32572
Identified using the e-PCR software (G. Schuler)"
146274..146413
/note="matching EMBL:F01603
RHdb:RH9897
dbSTS:STS25118
Identified using the e-PCR software (G. Schuler)"
183972..184175
/note="matching EMBL:T16335
RHdb:RH53640
dbSTS:STS42879
Identified using the e-PCR software (G. Schuler)"
183974..184130
/note="matching EMBL:T16335
RHdb:RH56383
dbSTS:STS46312
Identified using the e-PCR software (G. Schuler)"
54122 a 35054 c 36149 g 60833 t

ORIGIN

Query Match 3.9%; Score 42.8; DB 9; Length 186158;
Best Local Similarity 50.5%; Pred. No. 6.7; Mismatches 0; Gaps 0;
Matches 104; Conservative
QY 678 RAAGGAGACCAAGCTTCTAAGTAATCAAGAAAAATCTTGGACAATAGCTTTCCATA 737
Db 61470 AAAGACATATACAGATGCAATAAAGACATGAATAATGTTAAAAATCAITAGTATT 61529
QY 738 GAGAAAACCGTGAGGAAGAGTCGATACCATTAATTCGGATTAAAAAAGGAGCTGCTTA 797
Db 61530 AGGGAATGCAAGACATCATGAGATCAATTAAGTACCTGTCAGATGACTAAAAATAA 61589
QY 798 CAACAGCCATATCATCGGACCAACCTTCCTTTTGTGTTCTTTGTAATGACAGAGAAGTAA 857
Db 61590 AAAATGCCAAGATGTGGAACAACACAGCTCTTTTCATGATGCTTATGAGAATGTAATAAT 61649
QY 858 CAAGAGTAAGTTTTTTTGAATAATTTCT 883
Db 61650 GGTACAGCCATTCTAGGAACAGTTT 61675

RESULT 27
AC112384/c
LOCUS
Rattus norvegicus clone CH230-30L4, *** SEQUENCING IN PROGRESS ***
DEFINITION
65 unordered pieces.
AC112384
AC112384.3 GI:21738458
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus

REFERENCE
1 (bases 1 to 188385)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Edwards, C.C., Elhaj, C., Escotto, M., Farnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Foster, P., Frantz, P., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, W., Liu, W., Louised, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okuwon, G., Oraguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Paters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojloban, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G., and Gibbs, R.

TITLE

JOURNAL

AUTHORS

REFERENCE

JOURNAL

TITLE

JOURNAL

COMMENT

Unpublished
 2 (bases 1 to 188385)
 Worley, K.C.

Direct Submission
 Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 188385)
 Worley, K.C.

Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Jul 12, 2002 this sequence version replaced gi:20303170.

----- Genome Center of Medicine
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GRAP
 Center clone name: CH230-30L4

----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 129147 bases at least Q40
 Consensus quality: 135254 bases at least Q30
 Consensus quality: 139955 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: this is a 'working draft' sequence. It currently
 consists of 65 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 'N'. But the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1032: contig of 1032 bp in length
 1033: gap of unknown length
 1034: contig of 1659 bp in length
 2792: gap of unknown length
 2892: contig of 1123 bp in length
 4014: gap of unknown length
 4115: contig of 1075 bp in length
 5189: gap of unknown length
 5290: contig of 1696 bp in length
 6985: gap of unknown length
 7086: contig of 1666 bp in length
 8752: gap of unknown length
 8851: contig of 1125 bp in length
 9977: gap of unknown length
 10076: contig of 1076 bp in length
 12058: contig of 1982 bp in length
 12159: gap of unknown length
 13574: contig of 1416 bp in length
 13675: gap of unknown length
 14867: contig of 1193 bp in length
 14968: gap of unknown length
 16074: contig of 1107 bp in length
 16174: gap of unknown length
 16763: contig of 1499 bp in length
 16774: gap of unknown length
 17774: contig of 1089 bp in length
 18863: gap of unknown length
 20739: contig of 1777 bp in length
 20839: gap of unknown length

22054: contig of 1215 bp in length
 22154: gap of unknown length
 23276: contig of 1122 bp in length
 23277: gap of unknown length
 24790: contig of 1414 bp in length
 24800: gap of unknown length
 24931: contig of 1043 bp in length
 25933: gap of unknown length
 27512: contig of 1479 bp in length
 27612: gap of unknown length
 29296: contig of 1684 bp in length
 29396: gap of unknown length
 31289: contig of 1893 bp in length
 31389: gap of unknown length
 32807: contig of 1418 bp in length
 32907: gap of unknown length
 35198: contig of 2291 bp in length
 35298: gap of unknown length
 36909: contig of 1611 bp in length
 37009: gap of unknown length
 39110: contig of 2101 bp in length
 39210: gap of unknown length
 40983: contig of 1773 bp in length
 41084: gap of unknown length
 43481: contig of 2398 bp in length
 43581: gap of unknown length
 46034: contig of 2453 bp in length
 46134: gap of unknown length
 48083: contig of 1949 bp in length
 48183: gap of unknown length
 50384: contig of 2801 bp in length
 51084: gap of unknown length
 52750: contig of 1706 bp in length
 52850: gap of unknown length
 52891: contig of 2215 bp in length
 55105: gap of unknown length
 55205: gap of unknown length
 57621: contig of 2416 bp in length
 57721: gap of unknown length
 59974: contig of 2253 bp in length
 60074: gap of unknown length
 62471: contig of 2397 bp in length
 62571: gap of unknown length
 64507: contig of 1936 bp in length
 64607: gap of unknown length
 66410: contig of 1803 bp in length
 66510: gap of unknown length
 68760: contig of 2250 bp in length
 68860: gap of unknown length
 71981: contig of 3121 bp in length
 72081: gap of unknown length
 74426: contig of 2345 bp in length
 74526: gap of unknown length
 77301: contig of 2775 bp in length
 77401: gap of unknown length
 79708: contig of 2307 bp in length
 79808: gap of unknown length
 82290: contig of 2482 bp in length
 82390: gap of unknown length
 86138: contig of 3748 bp in length
 86238: gap of unknown length
 90775: contig of 4537 bp in length
 90875: gap of unknown length
 94169: contig of 3294 bp in length
 94269: gap of unknown length
 96988: contig of 2719 bp in length
 97088: gap of unknown length
 99272: contig of 2184 bp in length
 99372: gap of unknown length
 103549: contig of 4177 bp in length
 103649: gap of unknown length
 106784: contig of 3135 bp in length
 106884: gap of unknown length
 111169: contig of 4285 bp in length

```

* 111170 111269: gap of unknown length
* 111270 113821: contig of 2552 bp in length
* 113821 113921: gap of unknown length
* 113921 117375: contig of 3454 bp in length
* 117375 117475: gap of unknown length
* 117475 121282: contig of 3807 bp in length
* 121282 121282: contig of 3807 bp in length

Query Match      3.9%  Score 42.8; DB 2; Length 188385;
Best Local Similarity 49.5%; Pred. No. 6.7;
Matches 110; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 668 AAGAAATCAGAGAGAGACCGGAAAGCTTCTAAGTAATCAAGAAAAAATCTGGACAATAG 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171655 AAAAGAACTCAAGCAAGATGAACAAACACGAGAGATAACAAACAGAGAGATAAACT 171596

QY 728 CTTTTCATAGACAAACCGTCGAGAAAGAGTGCATACCATTAATTCGGATTAAAAAG 787
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171595 TTTATCCATACTAAACAGAGTCACAGAGTATCCAAATATCAATATCAAGAGAAAG 171536

QY 788 GGAGTGTTCACACGCCATATCAGTGGAGCAACCTCCCTTTTGATCTTTGATATGAC 847
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171535 GGAATATACACACAACTAAGGAATTCACAAAATATCAGATCCCTCTTCAAAATC 171476

QY 848 AGAAGATGACAAAGAGTAAGTTTGTGAAAATCTTAAATA 889
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171475 CTATGTTCAACAAACAGAAATTTGGATGAATGGACAAGA 171434

RESULT 28
AC092584/c
LOCUS      AC092584      159035 bp      DNA      linear      PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-20H7 from 2, complete sequence.
ACCESSION AC092584 AC015490
VERSION   AC092584.2 GI:16303525
KEYWORDS  HTG.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
            95063792
            9847074
REFERENCE  2 (bases 1 to 159035)
            Armstrong, J., Kozlowski, A., Hawkins, M. and Spalding, L.
            The sequence of Homo sapiens BAC clone RP11-20H7
            Unpublished (2001)
REFERENCE  3 (bases 1 to 159035)
            Waterston, R.H.
            Direct Submission
            Submitted (19-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 159035)
            Waterston, R.H.
            Direct Submission
            Submitted (20-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            5 (bases 1 to 159035)
            Waterston, R.H.
            Direct Submission
            Submitted (23-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            6 (bases 1 to 159035)
            Waterston, R.H.
            Direct Submission
            Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 20, 2001 this sequence version replaced gi:14916169.

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu

----- Summary Statistics
Center project name: H_NH0020H07
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-138C13, 2000 bp overlap; the clone sequenced to the right is RP11-21L7. Actual start of this clone is at base position 1 of RP11-20H7; actual end is at base position 159035 of RP11-20H7.

Data from AC068537, AC073994, and AC016827 was used to finish this clone, AC092584.

The sequence of AC015490 has been incorporated into AC092584.

FEATURES

source	Location/Qualifiers
1..159035	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-20H7"
	/clone_lib="RP11-11"
1..108	/rpt_family="L1"
85..149	/rpt_family="L1"
148..550	/rpt_family="L1"
344..387	/rpt_family="L1"
550..2451	/rpt_family="AT-rich"
2031..2068	/rpt_family="L1"
3780..4131	/rpt_family="(TG)n"
4308..4401	/rpt_family="MaLR"
	/rpt_family="L2"

repeat_region	5512..5561	/rpt_family="Mariner"
repeat_region	5545..5567	/rpt_family="AT-rich"
repeat_region	6515..7739	/rpt_family="L1"
repeat_region	7582..7622	/rpt_family="TCTA)n"
repeat_region	8664..8814	/rpt_family="MIR"
repeat_region	9632..9817	/rpt_family="L1"
repeat_region	9848..9990	/rpt_family="TCCA)n"
repeat_region	10104..10426	/rpt_family="L1"
repeat_region	10466..10564	/rpt_family="TTTTA)n"
repeat_region	10574..10605	/rpt_family="TTTTA)n"
repeat_region	10601..11026	/rpt_family="L1"
repeat_region	11062..11958	/rpt_family="L1"
repeat_region	11334..11362	/rpt_family="AT-rich"
repeat_region	11987..12289	/rpt_family="L1"
repeat_region	12777..12812	/rpt_family="AT-rich"
repeat_region	12799..13336	/rpt_family="L1"
repeat_region	13270..13291	/rpt_family="AT-rich"
repeat_region	13365..13877	/rpt_family="L1"
repeat_region	13879..14061	/rpt_family="G-rich"
repeat_region	14072..14232	/rpt_family="L1"
repeat_region	14231..14412	/rpt_family="L1"
repeat_region	14391..14436	/rpt_family="AT-rich"
repeat_region	14492..15506	/rpt_family="L1"
misc_feature	15546..15895	/note="similar to Homol (NID:g9808862)"
repeat_region	16673..16839	/rpt_family="Alu"
repeat_region	17085..18013	/rpt_family="L1"
repeat_region	18027..18380	/rpt_family="L1"
repeat_region	18381..18406	/rpt_family="T)n"
repeat_region	18684..16988	/rpt_family="L1"
repeat_region	19010..19060	/rpt_family="Mariner"
repeat_region	19069..19154	/rpt_family="L1"
repeat_region	19830..19852	/rpt_family="(GAAA)n"
repeat_region	21727..21845	/rpt_family="r-rich"
repeat_region	22856..22888	/rpt_family="TTTTC)n"
repeat_region	26258..26691	/rpt_family="L1"
repeat_region	27951..28074	/rpt_family="L1"

repeat_region	28081..29567	/rpt_family="I1"
repeat_region	29568..29848	/rpt_family="Alu"
repeat_region	29849..32251	/rpt_family="I1"
repeat_region	31201..31310	/rpt_family="T-rich"
misc_feature	32226..33096	/note="similar to Sus scrofa EST BI184750 (NID:g14659159);
misc_feature	32227..32588	
Query Match	3.9%	Score 42.4; DB 9; Length 159035;
Best local Similarity	52.8%	Prod. No. 8.5;
Matches 114; Conservative	0; Mismatches	101; Indels 1; Gaps 1;
QY 672	AATCAAGAAAGGACGCGAGACCTTCTAGTAATCAAGAAATAATCTTGACACAAATGCTT 731	
Db 21746	AAATCAATGAACAAAGAGTGTGTTTAAATTAACAAATTTGACAA-ACCTTTA 21688	
QY 732	TCCATAGAGAAACCGTGAGGAAGGAGTCGATACCATTAATTCGGATTAAGAAAGGAG 791	
Db 21687	GCCAGACAAAGAAAATAAGAAAGGTCGCAATAATAATTAATCAAGATGAATGAG 21628	
QY 792	TGCTTACACAGCCATATCAGTGGAGACCTTCCCTTTGATCTTTTGATATGACAGAA 851	
Db 21627	AGCTTACAACCTGCTATCCAGAAATTCAGAGGATATTTGCGCTCTATGACGACGTA 21568	
QY 852	GATGAACAAAGAGTAAGTTTTTCAGAAAATCTCTAAAA 887	
Db 21567	AGTCACAAAGTGAAAATCTAGAGAAATGTTAAA 21532	
RESULT 29		
AF268060		
LOCUS	AF268060	2706 bp DNA linear BCT 24-MAY-2000
DEFINITION	Candidatus Carsonella ruddii natural-host Aphalaroida inermis RNA polymerase beta subunit (rpoB) and RNA polymerase beta-prime subunit (rpoC) genes, partial cds.	
ACCESSION	AF268060	
KEYWORDS	AF268060.1	GI:14193385
SOURCE	Candidatus Carsonella ruddii.	
ORGANISM	Candidatus Carsonella ruddii.	
REFERENCE	1. (bases 1 to 2706)	
AUTHORS	Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.	
TITLE	Phylogenetic analysis of vertically transmitted psyllid endosymbionts (Candidatus Carsonella ruddii) based on atpAGD and rpoC: comparisons with 16S-23S rDNA-derived phylogeny	
JOURNAL	Unpublished	
REFERENCE	2. (bases 1 to 2706)	
AUTHORS	Thao,M.L., Clark,M.A., Burckhardt,D.H., Moran,N.A. and Baumann,P.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-MAY-2000) Microbiology Section, University of California, Davis, One Shields Avenue, Davis, CA 95616-8665, USA	
FEATURES	Location/Qualifiers	
source	1..2706	
	/organism="Candidatus Carsonella ruddii"	
	/specific_host="Aphalaroida inermis"	
	/db_xref="taxon:114186"	
	<1..184	
	/gene="rpoB"	
	<1..184	
	/gene="rpoB"	
	/codon_start=2	
	/transl_table=11	
	/product="RNA polymerase beta subunit"	
	/protein_id="AAK55948.1"	
	/db_xref="GI:14193386"	
	/translations="YGAANILREMLTIKSDDLGRCTCFKNIINDIHYINFGIPESF	
	ILIQEIKSLCFDIIV"	
	181..2706	
gene		
CDS		

CDS /gene="rpoC"
181..>2706
/gene="rpoC"
/codon_start=1
/transl_table=1
/product="RNA polymerase beta-prime subunit"
/protein_id="AAK55949.1"
/db_xref="GI:14193387"
/translation="MKKIKLIIKASPKNKIKWSYGEIKNSFNFKNLKPEKNGLF
CKLIFSYNICEKKKCNKNNKFLNLRKSHSYRGHTKLYPAIHWFYKNNNT
SLISLSEKILKIKIVLSEKPKKYSLSLKNILNLCYVLSGAKALKML
SNEILLDCVLLKLLKNCSENLFFYLDKTKIYFVLSGNKPEWFLIPIVSP
KTRPLIPNIGKYSALDNLRYKKVDNRILKKFKILKQIINERILLQESINA
LFDNKEIENPLTNSKRVLSFSNSIKGYIFRQNLKGRVDFSGRSVSYDPELL
YCKPILIGIEIFKPIFNKAKKIIENINFLDEFYLNKKKESINILKIDCKMKT
LNLRAPLHRMGISFKILTQDKTKIHPILCLSYNADFQDQNAIHLPIITINAQVE
SNYLLSNNISPSNGEPIIIPQDVIYGLVCLTNNNKILPSSFFDVKNFNK
NKAENVLNKNYFIKTTVGRIFYYILKKTINFNKIKKDLIYIKYIFEYF
GVYKTKILDNKKIGFFFTFFGTFISYDLPKIKENCLKLIKTKNNKDNFINI
NVVELLIENIIKIKITNNKNNLFIIMDSGRSILQIKLLGPRGFSKNGEI
IIDPLDNKGLSKMNYFISFGARKGLTDTSLKANSGLYLRKLVDMOGIVYKI
DCGVYGEILIKNNNFYLLIKIFGRITIDISIFPKNLIKIKNTLNKNIIPLIK
KKINKIYIRSVLHCVSIRGICVLCYGDLSNKLINIGTPGII"
BASE COUNT 1199 a 188 c 266 g 1053 t
ORIGIN

Query Match 3.9%; Score 42.2; DB 1; Length 2706;
Best Local Similarity 50.2%; Pred. No. 15;
Matches 104; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 830 TGATTCCTTTGATACAGAGAAGTCAACAGAGTAAAGTCTTAAATA 889
DB 2124 TAATTTATTAATGTTAGTCTGGTCAAGAGAGTATTTTACAATCAACAAATT 2183
QY 890 GAAATGATTTCTGCTGACCTAAAGCTTCTCTCTTTATTTCCITTTGCTAGATGATT 949
DB 2184 ATTAGTTTATAGAGGATTTTTCCTAAATCTAATGGGAATATTTATGATCCCAATTT 2243
QY 950 CGGTAGTAGGAATATGGGTGAGATGACGCTGGAGTGAATGCGTTCGTGGAAGAAC 1009
DB 2244 AGATATTTTAAATAGGATTAAGTATGAAAATATTTTAAAGCTCTTTGGAGCAAG 2303
QY 1010 AGGGAAGTATGCTCTGACGCTTTAGA 1036
DB 2304 AAAAGGATTAACGTACTCTTTTAAA 2330

RESULT 30
AC1010342/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2015B23, linear PRI 01-MAY-2001
ACCESSION AC1010342
VERSION AC1010342.5 GI:13876487
KEYWORDS HTG
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 126010)
DOE Joint Genome Institute.
Direct Submission
2 (bases 1 to 126010)
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 126010)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 126010)
Submitted (28-APR-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 126010)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
TITLE Submitted (01-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 28, 2001 this sequence version replaced gi:13699503.
Draft sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content
SHGC-104285 G58244
WI-4392 G02894.
Location/Qualifiers
1. 126010
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2015B23"
BASE COUNT 41323 a 21680 c 22194 g 40813 t
ORIGIN

Query Match 3.9%; Score 42.2; DB 9; Length 126010;
Best Local Similarity 43.6%; Pred. No. 9; 7;
Matches 188; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 465 AAGGCACGTGCGCTTAAAGGTCATGGATGAGATCAGGAGTTAAGGAGAAATATCTTTAT 524
DB 56662 AGAATTTTGCATAACATAAACATGTTTGCATGATGATTTGTCACCTGCCCTCA 56603
QY 525 GAGAGTTTCGCATTACACAGCGCAATGGGAAAGTAAAGGTCCTGCAACAGAGACAAGTATT 584
DB 56602 AAAGATTAACAGGAAATAGCAATTTAAATATATGCGGCACATCTACAGAGCAACAACT 56543
QY 585 GATGGGAGCTACAAATCCCTGAAACGGGATATAGAGAAATATGCTTTCTTAATTCGAAGA 644
DB 56542 TAATAGATAACAAATTTACTCAAAATTTTAAATATCAATATATATATTTATTTGCTAAGA 56483
QY 645 ATTTAATGACCTGCTTGATGCTTAAGAAATCAGAAAGGAGCCGAAAGGCTTCTAAGTAAT 704
DB 56482 AAGTCATTTGATTAGAAATTTGAAAGATAGTGTATATATATACCAATAAATTTGGAAT 56423
QY 705 CAAGAAAAATCTTTGGCAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGAGTGCAT 764
DB 56422 TTAATTAAGGCGCTTAAATAACTTTTGTGTTTAAAGATATGTTTACAAAGGAAATATGG 56363
QY 765 ACCATAAATTCGGATTTAAAGGAGGAGTCTTACAAACAGCCCATATATCAGTGAGCAACCTT 824
DB 56362 CCTCAACTGAAGGACATTAAGTATGCTATATAAAACCATGAGATACCTGAGTATAGTA 56303
QY 825 CCTTTGATCTTTTGAATATGACAGAGATGAACAAAGAGTAAGTTTTTGAATAATTTCTA 884
DB 56302 CACATGGAGGTATTATTACAGCCTTAACCCATCGCTTCAGAAAACGTTTGAANAATCACTA 56243
QY 885 AAATAGAAATG 895
DB 56242 AGTTTGAATG 56232

RESULT 31
AC108007/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone CTD-3038L12 map 11, *** SEQUENCING
IN PROGRESS ***, 8 ordered pieces.
ACCESSION AC108007
VERSION AC108007.4 GI:22474954
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141635)

AUTHORS JOURNAL REFERENCE AUTHORS

Birren, B., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 11, clone CTD-3038L12
Unpublished
2 (bases 1 to 141635)
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B.,
Brown, A., Camarata, J., Chang, J., Chazaro, B.,
Cooke, P., DeArelano, K., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Farro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
Mlenda, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 141635)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Farro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Macdonald, P., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Meldrim, J., Meneus, L., Mihova, T., Mlenda, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 24, 2002 this sequence version replaced gi:20336822.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Genome Center
Center code: WBIR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L24596
Center clone name: 3038_L12

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 92004: contig of 92004 bp in length
* 92005 92104: gap of 100 bp
* 92105 117227: contig of 25123 bp in length
* 117228 117327: gap of 100 bp
* 117328 120434: contig of 3107 bp in length
* 120435 120534: gap of 100 bp
* 120535 126202: contig of 5668 bp in length
* 126203 126302: gap of 100 bp
* 126303 129158: contig of 2856 bp in length
* 129159 129258: gap of 100 bp
* 129259 131961: contig of 2703 bp in length
* 131962 132061: gap of 100 bp
* 132062 135739: contig of 3678 bp in length
* 135740 135836: gap of 100 bp
* 135840 141635: contig of 5796 bp in length.

FEATURES source

1. 141635
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="CTD-3038L12"
/clone_lib="CTD2 Human BAC"
BASE COUNT 42227 a 27391 c 27480 g 43740 t 797 others
ORIGIN

Query Match 3.9%; Score 42.2; DB 2; Length 141635;
Best Local Similarity 43.6%; Pred. No. 9.6;
Matches 188; Conservative

QY 465 AAGGACGCTGCCCTAAAGGTCATGATGAGATCAGGAGTTAAGGAGAAATATCTTTAT 524
DB 119252 AAGAAATTTGCATAACATAACATGTTTGACATGATGAATATCTTGCACCTCTGCCCTCA 119193
QY 525 CGAGAGTTTCGCTTACAGAGGCAATGGGAAAGTACCGCTCTGCAACAGAGAGCAAGTAAT 584
DB 119192 AAGAGATACAGGAAATAGCAATTTAAAATATATACGCACATCTCAGAGCAACAACT 119133
QY 585 GATGGGAGCTAGCAATCTGAAACGGGATATAGGAAATATGTTTCTTAATTCGAAGA 644
DB 119132 TAATACATACAAATTTACTCAAAATTTAAATATCAATATATATATATATATATAT 119073
QY 645 ATTTAATGACTGCTTGATGCTCAAGAAATCAGAGGAGACCGGAAAGCTTCTAAGTAAT 704
DB 119072 AAGTCAATTTAGTAATTTAGAGATAGTGTATATATATATATATATATATATATAT 119013
QY 705 CAGAGAAATCTTGGACATAGCTTTTCCATAGAGAAACCGTCAGAGGAGGAGTCGAT 764
DB 119012 TTAATAAAGCCCTACTAAATACTTTTGGTTTAAAGATATGTTTACAAAGGAATATGG 118953
QY 765 ACCATAAATTCGATTTAAAGAGGAGTGTCTACACAGCCATATCTAGTGGAGCACTT 824
DB 118952 CCTCAACTGAAGGACATTAAGTATGCTATATAAAACCATGAGATACCTGAGTATAGTA 118893
QY 825 CCCTTTGATTTCTTTGATTAATCAGAGAGATGAACAAGAGTAGTAAGTTTTTGAAAAATCTA 884
DB 118892 CACATGGAGTATTATTACAGCCCTAAACCCATCGCTTTCAGAAACGTTTGAAAAATCACTA 118833
QY 885 AATAGAAATG 895
DB 118832 AGTTTGAATG 118822

RESULT 32
AL807375/c 192168 bp DNA linear HTG 17-JUL-2002
LOCUS Mus musculus chromosome X clone RP23-247H23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 8 unordered pieces.
AL807375
ACCESSION AL807375.5 GI:21738714
VERSION HTG; HTGS_PHASE1.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 92824: contig of 92824 bp in length
 92825 92924: gap of 100 bp
 92925 102534: contig of 9610 bp in length
 102535 102634: gap of 100 bp
 102635 109946: contig of 7312 bp in length
 109947 110046: gap of 100 bp
 110047 112095: contig of 2049 bp in length
 112096 112195: gap of 100 bp
 112196 202801: contig of 90606 bp in length.

FEATURES

Source
 1. .202801
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-146O20"
 /clone_lib="RPCI-23"
 1. 92824
 /note="assembly fragment:01268
 fragment_chain:1"
 92825..102534
 /note="assembly fragment:01238
 fragment_chain:1"
 102635..109946
 /note="assembly fragment:03032
 fragment_chain:1"
 110047..112095
 /note="assembly fragment:01164
 fragment_chain:1"
 112196..202801
 /note="assembly fragment:00215
 fragment_chain:1"
 64339 a 38445 c 38425 g 61192 t 400 others

BASE COUNT 64339 a 38445 c 38425 g 61192 t 400 others
 ORIGIN

Query Match 3.9%; Score 42.2; DB 2; Length 202801;
 Best Local Similarity 55.0%; Prod. No. 9.2;
 Matches 83; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 188 AATTCCTAAAGGAGAACACGCTACCTTATCAATGAAAGCGTGCCCAAGTCTCGAGGAGAA 247
 Db 190083 AATGTTGAAGCAGAGAAATCTCAATTAACATTAAGCAAGGCTCAAGTCTCGAGAGAA 190142
 QY 248 CTGATAGAACGCTCCATATGCTTTTCTTCCTTCATGCTTTTCAATAAGTCCGCTTCA 307
 Db 190143 CACATAGTTACTTCTAGGAATCAATTTTTCCTTCATGGAATAAGGCGATATGCTTTAA 190202
 QY 308 ACAGTAACCAAGCAGCTACATAGCGA 338
 Db 190203 AATGAACCTTACCTCGCTTCATTAGAGA 190233

RESULT 34
 AP000988/c 266050 bp DNA linear BCT 11-OCT-2001
 LOCUS
 DEFINITION Sulfolobus tokodaii genomic DNA, complete sequence, section:8/10.
 ACCESSION AP000988 BA000023
 VERSION AP000988.1 GI:15622956
 KEYWORDS

SOURCE Sulfolobus tokodaii (strain:7) DNA.
 ORGANISM Sulfolobus tokodaii
 Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

Sulfolobus.

REFERENCE
 AUTHORS
 1 Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M.,
 Sekine, S., Baba, S., Anai, A., Kosugi, H., Hosoyama, A., Fukui, S.,
 Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M.,
 Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J.,
 Kushida, N., Oouchi, A., Aoki, K., Masuda, S., Yanagii, M.,
 Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H.

TITLE

Complete genome sequence of an aerobic thermoacidophilic
 Crenarchaeon, Sulfolobus tokodaii strain7
 DNA Res. 8 (4), 123-140 (2001)

REFERENCE

2 (bases 1 to 266050)
 Kawarabayasi, Y., Tanaka, T., Hino, Y. and Kikuchi, H.

AUTHORS

Direct Submission

TITLE

Submitted (22-DEC-1999) Yutaka Kawarabayasi, National Institute of
 Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10,
 Shibuya-ku, Tokyo 151-0066, Japan (E-mail: kyutaka@nite.go.jp,
 URL: http://www.bio.nite.go.jp/; Tel: 81-3-3481-8951,
 Fax: 81-3-3481-8424)

COMMENT

Kawarabayasi, Y. is officially affiliated with the National
 Institute of Advanced Industrial Science and Technology, Tsukuba,
 Ibaraki 305-0046, Japan
 Yamagishi, A. and Oshima, T. are at Tokyo University of Pharmacy
 and Life Science, Hachioji, Tokyo 192-0392, Japan
 The other authors are at the National Institute of Technology and
 Evaluation, Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan
 Homology analysis is performed by Smith-Waterman algorithm against
 Genbank release 109; EMBL release 36.0; SwissProt release 38.0;
 PIR-Protein release 62.0; and OML release 31.4.
 E-mail address for comments and questions: kyutaka@nite.go.jp
 ORF organization, sequence alignment and more information are
 available at W.W.W. site of Biotechnology Center,
 URL: http://www.bio.nite.go.jp/dogan/genome_list-e.html/.

FEATURES

source
 1. .266050
 /organism="Sulfolobus tokodaii"
 /strain="7"
 /db_xref="taxon:111955"
 288..830
 /gene="ST1855"
 288..830
 /gene="ST1855"
 /note="OML:AAD0XP24H1 percent identity: 83.432 in 169aa."
 /codon_start=1
 /transl_table=11
 /product="180aa long hypothetical protein"
 /protein_id="BAB66948.1"
 /db_xref="GI:15622957"
 PNSSVFGKLVNFDHPAGPKPPLIATLENHAFDVFETLFTSYETIAGLFLVIGLL
 TRLAAGAAVMAIGMAPYAWLGSCTDEBQWIGSLITAGAIWVMTGAGRWGLDYFLY
 824..1342
 /gene="ST1856"
 824..1342
 /gene="ST1856"
 /note="OML:AAD0XP24H2 percent identity: 68.590 in 156aa."
 /codon_start=1
 /transl_table=11
 /product="172aa long hypothetical protein"
 /protein_id="BAB66947.1"
 /db_xref="GI:15622958"
 /translation="MVRKMKVAIVGIIFALFTVWLGITGQWYNNVGVPLNNSKL
 PKLVITVITANTPQGLCLIMNITDIDGPAYAPAPIMEISNGTWHIELSSQIAND
 TVKTIQAPWNAKKSVNNYSGFVILGSEGFHLIKGLHLSPGKRYEVLKLYTPAISS
 NRQAIAYFTISS"
 1378..2772
 /gene="ST1857"
 1378..2772
 /gene="ST1857"
 /note="PIR:D64486 percent identity: 34.286 in 105aa."
 /codon_start=1
 /transl_table=11
 /product="464aa long conserved hypothetical protein"
 /protein_id="BAB66948.1"
 /db_xref="GI:15622959"
 /translation="MENYSKEITPERASQTSYDGGTGLGKRGKNIIGIVKIDIVRVE
 IDLSKGRCCRCRCNGCHLYAIKMSYKFGYIDFDSLENKITEINKRELGLIVTL
 IEKFPFIANYLIPIENAKISLEBYINLIKONGPENISFTDFLINNRKINKDDIFI
 ILDTIASCKSKCFYNFITERFDENLMKTLNILLKEKVEDDKIKLEKIKDKYGN

gene

CDS

AUTHORS
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE
2 (bases 1 to 98697)
Ozanich,A. and Mead,K.
The sequence of Homo sapiens PAC clone RP4-673M15
Unpublished (2001).

AUTHORS
JOURNAL
3 (bases 1 to 98697)
Waterston,R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
4 (bases 1 to 98697)
Waterston,R.H.
Direct Submission
Submitted (14-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

AUTHORS
JOURNAL
5 (bases 1 to 98697)
Waterston,R.
Direct Submission
Submitted (22-OCT-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

REFERENCE
6 (bases 1 to 98697)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

AUTHORS
JOURNAL
7 (bases 1 to 98697)
Waterston,R.
Direct Submission
Submitted (04-JUN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 4, 2002 this sequence version replaced gi:4827328.

COMMENT

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu

Contact: Summary Statistics

Center project name: H_DJ0673M15

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GB/CHR7>, send mailto:regreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994).

The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCIPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is DJ0647J21. Actual start of this clone is at base position 1 of DJ0673M15.

Base positions 19689 to 19719 of the submitted sequence is represented by single chemistry only.

FEATURES	source
Location/Qualifiers	1..98697
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="7"	
/map="7p13-p11.2"	
/clone="RP4-673M15"	
/clone.lib="RPCI-4"	
4..568	
/note="similar to EST AW131540 (NID:g6133147) xf31b08.x1"	
4..568	
/note="similar to EST BF026668 (NID:g10734380)"	
4..548	
/note="similar to EST AA775847 (NID:g2835181) ad17e01.s1"	
4..540	
/note="similar to EST BE727822 (NID:g10141814)"	
4..318	
/note="similar to EST AI017276 (NID:g3231612) ov27f07.x1"	
4..310	
/note="similar to EST AA216094 (NID:g1816041)"	
33..552	
/note="similar to EST BE075304 (NID:g8424032)"	
177..568	
/note="similar to EST BE588888 (NID:g9841927)"	
216..568	
/note="similar to EST BE256638 (NID:g9127101)"	
254..568	
/note="similar to EST AA057332 (NID:g1550170) zf56all.r1"	
554..857	
/rpt_family="Alu"	
862..1003	
/note="similar to EST BE727822 (NID:g10141814)"	
886..1080	
/note="similar to EST AA057332 (NID:g1550170) zf56all.r1"	
886..1080	
/note="similar to EST BF026668 (NID:g10734380)"	
886..1079	
/note="similar to EST BE588888 (NID:g9841927)"	
886..1075	
/note="similar to EST BE256638 (NID:g9127101)"	
886..922	
/note="similar to EST AW131540 (NID:g6133147) xf31b08.x1"	
1381..1662	
/rpt_family="L1"	
1663..1954	
/rpt_family="Alu"	
1673..1727	
/note="similar to EST A1147019 (NID:g3674701) qs82a12.x1"	
1959..2254	
/rpt_family="Alu"	
2255..2303	
/rpt_family="L1"	
2329..2467	
/rpt_family="L1"	
2468..2522	
/rpt_family="L1"	
2912..3207	
/rpt_family="Alu"	
3239..3404	
/rpt_family="Alu"	
3625..3935	


```

----- Summary Statistics
Sequencing vector: Plasmid;
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 59885 bases at least Q40
Consensus quality: 63135 bases at least Q30
Consensus quality: 65093 bases at least Q20
-----
** NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
** NOTE: This is a 'working draft' sequence. It currently
** consists of 69 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.
**
1125: contig of 1125 bp in length
1126: gap of unknown length
1246: contig of 1020 bp in length
2345: gap of unknown length
3896: contig of 1551 bp in length
3897: gap of unknown length
5572: contig of 1576 bp in length
5673: gap of unknown length
6698: contig of 1026 bp in length
6799: gap of unknown length
8141: contig of 1343 bp in length
8142: gap of unknown length
9510: contig of 1269 bp in length
9511: gap of unknown length
9610: contig of 1495 bp in length
11105: contig of 1495 bp in length
11205: gap of unknown length
12766: contig of 1561 bp in length
12767: gap of unknown length
14715: contig of 1849 bp in length
14716: gap of unknown length
14815: contig of 1058 bp in length
15873: gap of unknown length
15974: contig of 1248 bp in length
17221: contig of 1124 bp in length
17322: gap of unknown length
18445: contig of 1124 bp in length
18446: gap of unknown length
19828: contig of 1283 bp in length
21841: contig of 1913 bp in length
21842: gap of unknown length
23280: contig of 1339 bp in length
23281: gap of unknown length
23282: contig of 1841 bp in length
23283: gap of unknown length
23284: contig of 1588 bp in length
25222: contig of 1588 bp in length
25223: gap of unknown length
26910: gap of unknown length
27010: contig of 1000 bp in length
28109: gap of unknown length
28110: contig of 1675 bp in length
29784: gap of unknown length
29885: contig of 1666 bp in length
31550: gap of unknown length
31551: contig of 1510 bp in length
33160: gap of unknown length
33260: contig of 1412 bp in length
34772: gap of unknown length
34773: contig of 1485 bp in length
36257: gap of unknown length
36358: contig of 1051 bp in length
37408: contig of 1051 bp in length
37508: gap of unknown length
39418: contig of 1910 bp in length
39518: gap of unknown length
40833: contig of 1315 bp in length
40934: gap of unknown length
42346: contig of 1413 bp in length
42347: gap of unknown length
42447: contig of 1271 bp in length
43117: gap of unknown length
43118: contig of 1186 bp in length
45003: gap of unknown length
45103: contig of 1929 bp in length
45104: gap of unknown length
47132: contig of 1319 bp in length
47133: gap of unknown length
48551: gap of unknown length
48552: contig of 2252 bp in length
49503: gap of unknown length
50493: contig of 1146 bp in length
52049: gap of unknown length
52149: contig of 1338 bp in length
53487: gap of unknown length
53587: contig of 1278 bp in length
54865: gap of unknown length
54866: contig of 1927 bp in length
56992: gap of unknown length
56993: contig of 1476 bp in length
58468: gap of unknown length
58469: contig of 1036 bp in length
59604: gap of unknown length
59605: contig of 1728 bp in length
61432: gap of unknown length
61433: contig of 1893 bp in length
63425: gap of unknown length
63426: contig of 1693 bp in length
65218: gap of unknown length
65219: contig of 1444 bp in length
66762: gap of unknown length
66763: contig of 1224 bp in length
68086: gap of unknown length
68087: contig of 1371 bp in length
69557: gap of unknown length
69558: contig of 1886 bp in length
71543: gap of unknown length
71544: contig of 1163 bp in length
72806: gap of unknown length
72807: contig of 1705 bp in length
74611: gap of unknown length
74612: contig of 1831 bp in length
76542: gap of unknown length
76543: contig of 2180 bp in length
78822: gap of unknown length
78823: contig of 1095 bp in length
80017: gap of unknown length
80018: contig of 2589 bp in length
82706: gap of unknown length
82707: contig of 2207 bp in length
85013: gap of unknown length
85014: contig of 1611 bp in length
85114: gap of unknown length
85115: contig of 1611 bp in length

```

Query Match 3.9% Score 42; DB 2; Length 124767;

Best Local Similarity 47.4%; Pred. No. 11;

Matches 126; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```

Qy 663 GGCTAAGAAATCAGAAAGGAGAGCCGAAAGCTTCTAAGTAATCAAGAAAAATCTCGAC 722
Db 95477 GGTCATGAATCTGGAGTGTAAAAATAAATAAGTAATAACCAACCAACCAATTAAT 95418

Qy 723 AATAGCTTTTCCATAGAGAAACCGTGAGGAAGAGTGCATACCAATAAAATCCGATTAA 782
Db 95417 AATAAATAAATAAATTCCTCAAGATTTATCCCAATGAGAAGATTAATAAATTAAGATTA 95358

Qy 783 AAAGGAGTGTCTACACAGCAGATATCAGTGGAGCAACCTCCCTTTGATTCTTGATA 842
Db 95357 AATAAATAAATCTCTTAAGCCAAACATTAAGTCCCTGTTCTTAGTTGATGATA 95298

Qy 843 ATGACAGAGATGAACAAAGAGTAGTTTTTGAAGAAAAATCTTAAGTAAGATGATTTGT 902
Db 95297 ACAGCAAAATTTGTTTTTTTCTACCTTCACATGACAGATACAAAGCCGATGAGTGTGT 95238

```

QY 903 GTCGAGCTAAAGCTGCTCTCTTTA 928
 Db 95237 CCCAAATACAGCTTGTCTTAGTTGA 95212

RESULT 37
 MC445/c
 LOCUS
 DEFINITION M.caprlicolom DNA for CONTIG MC445.
 ACCESSION 233336
 VERSION 233336.1 GI:516259
 KEYWORDS adenylsuccinate synthetase.
 SOURCE mycoplasma capricolom.
 ORGANISM Bacteria: Firmicutes; Mollicutes; Entomoplasmatales; Entomoplasmataceae.
 1 (bases 1 to 363)
 Bork.P., Ouzounis.C., Casari.G., Sander.C., Dolan.M. and Gillevet.P.
 Exploring the Mycoplasma capricolom genome: A Parasite reveals it's physiology
 Unpublished
 2 (bases 1 to 363)
 Gillevet.P., Barton.F., Brenner.S.E., Clark-Whitehead.R., Dolan.M., Douglas.N., Hsu.E., Purzcki.M.S., Richter.B., Russo.S., Sartell.J., Smith.S.W., Wang.C., Williams.J. and Gilbert.W.
 Direct Submission
 Submitted (13-JUL-1994) Mycoplasma capricolom Genome Project, Harvard University, Cambridge, MA 02138 Prokaryotic Genomes Project
 Institute for Computational Sciences and Informatics George Mason University, Fairfax, Virginia, USA, 22030-4444 E-mail: gilleve@euranus.ncgr.nih.gov
 3 (bases 1 to 363)
 Bork.P., Ouzounis.C., Casari.G., Schneider.R., Sander.C., Dolan.M., Gilbert.W. and Gillevet.P.M.
 Exploring the Mycoplasma capricolom genome: a minimal cell reveals its Physiology
 Mol Microbiol. 16 (5), 935-967 (1995)
 96059641
 7476192

NOTES:
 The tga codon is read as Tryptophan in Mycoplasma capricolom Coding sequences below were putatively identified by Peer Bork using the program Genequizz at the EMBL.
 EMBL, Meyerhofstr.1, 69012 Heidelberg, Germany
 E-mail bork@embl-heidelberg.de
 More than 870,477 bases have been sequenced by the genomic walking technique and assembled into 372 non-overlapping contigs (Accession numbers Z33005 to Z33376) covering 214,528 base pairs. The length of the contigs varies from 63 base pairs to 2049 base pairs with an average length of 658 base pairs. The data consists of 13,091 bases (6%) from one pass fluorescent sequencing and 201,437 bases (94%) produced by genomic walking. Of the latter, 155,805 bases (77%) have multiple coverage on at least one strand and 112,621 bases (56%) are covered on both strands. There is a total of 968 ambiguous nucleotide assignments in the data set (0.5%). We have compared 8,868 bases of our data with Mycoplasma capricolom sequences already stored in the public databases and note less than 0.7% difference between the two data sets (including ambiguous calls, insertions, deletions and mismatches).
 Location/Qualifiers
 1..363
 /organism="Mycoplasma capricolom"
 /strain="ATCC 73433(Kid)"
 /db_xref="taxon:2095"
 complement(1..>363)
 /note="ORF identified by homology to SwissProt entry swiss|P30520|PURA_HUMAN; Probability 3.8-33"
 /codon_start=3
 /evidence=not_experimental
 /transl_table=4
 /codon=(seq:"tga",aa:Ttp)
 /product="adenylsuccinate synthetase"

FEATURES
 source

CDS

/protein_id="CAA83842.1"
 /db_xref="GI:530493"
 /db_xref="SPTREMBL:Q49087"
 /translation="YFNNWNSCKTSKIVKYNTRVSGAMPTKTEIKELANKLR
 ERGREGYNTGKPRRIGWLDLVALKYAIRVGGIDQLFLFDVLDTEBKIKICTAYKL
 DNOIHSIPANENFDKV"
 BASE COUNT 112 a 64 c 36 g 150 t 1 others
 ORIGIN

Query Match 3.8%; Score 41.8; DB 1; Length 363;
 Best Local Similarity 48.7%; Pred. No. 23;
 Matches 146; Conservative 0; Mismatches 148; Indels 6; Gaps 1;

QY 606 AAACGGGATATAGGAAAAATGTTTCTTAATTCGAAGAATAATTAAGCTGCTTGATGCT 665
 Db 333 AAACAATTCAAAAGTCATAGGATGTTTAAAGCATATATAGTAGAGTTGGTTAGGAG 274
 QY 666 CTAAGAATCAGAAGAGAGACCAAGCTTCTAAGTAATCAAGAAAAAATCTTGGACAAT 725
 Db 273 CTATGCCAACNGAATTTAAACTTGAACCTTGCCTAATAACTAAGAGAACGTTGGAGAGAAT 214
 QY 726 AGCTTTCCATAGAGAAAACCGTGAGGAAGAGTCGATACCATAAATCCGATTAAAAA 785
 Db 213 ATGTTCTTAATACAGGAAGAACCAAGAGAAATAGGTTGATTAGATTTCGCTTTAAAAAT 154
 QY 785 AGGGAGTGCTTTACAAACAGCCATATACATAGTCGAGCAACCTTCCTTTGATTCTTTGATAATG 845
 Db 153 ATGCAATTT-----AGAGTTGGTGGAAATTTGATCAATTTATTTAACTTTATTTGATCTAT 100
 QY 846 ACAGAAGATGAACAAAGAGTAAGCTTTTTCAAAAATTCCTAAAAATAGAATGCATTTGTGTC 905
 Db 99 TAGATACAGAAGAAAAAATTTAAATTTGTACAGCTTATAAATAGATAATCAATATTATTC 40

RESULT 38
 AC024393/c
 LOCUS
 DEFINITION Homo sapiens chromosome 1 clone RP11-711C8 map 1, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC024393.1 GI:7108190
 VERSION AC024393.1
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 76141)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 1, clone RP11-711C8
 Unpublished
 2 (bases 1 to 76141)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Fenesor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,R., Jones,C., Kann,L., Karatas,A., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Hearford,A., Horton,L., Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenda,V., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivat,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission

TITLE

Fri Jan 10 12:01:26 2003

us-09-438-185a-l_1199590_1200675.rge

Page 43

JOURNAL
COMMENT

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIDR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6722

Center clone name: 711_C_8

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 736 835: contig of 735 bp in length
* 836 835: gap of 100 bp
* 1596 1695: contig of 760 bp in length
* 1596 1695: gap of 100 bp
* 2462 2561: contig of 766 bp in length
* 2462 2561: gap of 100 bp
* 2562 3364: contig of 803 bp in length
* 3365 3464: gap of 100 bp
* 3465 4249: contig of 785 bp in length
* 4250 4349: gap of 100 bp
* 4350 5141: contig of 792 bp in length
* 5142 5241: gap of 100 bp
* 5242 6017: contig of 776 bp in length
* 6018 6117: contig of 100 bp
* 6118 6900: contig of 783 bp in length
* 6901 7000: gap of 100 bp
* 7001 7784: contig of 784 bp in length
* 7785 7884: gap of 100 bp
* 7885 8666: contig of 782 bp in length
* 8667 8766: gap of 100 bp
* 8767 9543: contig of 776 bp in length
* 9543 9642: gap of 100 bp
* 9643 10391: contig of 749 bp in length
* 10392 10491: gap of 100 bp
* 10492 11278: contig of 787 bp in length
* 11279 11378: gap of 100 bp
* 11379 12148: contig of 770 bp in length
* 12149 12248: gap of 100 bp
* 12249 13033: contig of 785 bp in length
* 13034 13133: gap of 100 bp
* 13134 13889: contig of 756 bp in length
* 13890 13989: gap of 100 bp
* 14815 14914: contig of 825 bp in length
* 14915 15682: contig of 768 bp in length
* 15683 15782: gap of 100 bp
* 15783 16561: contig of 779 bp in length
* 16562 16661: gap of 100 bp
* 16662 17452: contig of 791 bp in length
* 17453 17552: gap of 100 bp
* 17553 18316: contig of 764 bp in length
* 18317 18416: gap of 100 bp
* 18417 19192: contig of 776 bp in length
* 19193 19292: gap of 100 bp
* 19293 20055: contig of 763 bp in length
* 20056 20155: gap of 100 bp
* 20156 20919: contig of 764 bp in length
* 20920 21019: gap of 100 bp

* 21020 21785: contig of 766 bp in length
* 21786 21885: gap of 100 bp
* 21886 22661: contig of 776 bp in length
* 22662 22761: gap of 100 bp
* 22762 23543: contig of 782 bp in length
* 23544 23643: gap of 100 bp
* 23644 24426: contig of 783 bp in length
* 24427 24526: gap of 100 bp
* 24527 25297: contig of 771 bp in length
* 25298 25397: gap of 100 bp
* 25398 26198: contig of 799 bp in length
* 26197 26296: gap of 100 bp
* 26297 27086: contig of 790 bp in length
* 27087 27186: gap of 100 bp
* 27187 27978: contig of 792 bp in length
* 27979 28078: gap of 100 bp
* 28079 28864: contig of 786 bp in length
* 28865 28964: gap of 100 bp
* 28965 29718: contig of 754 bp in length
* 29719 29818: gap of 100 bp
* 29819 30601: contig of 783 bp in length
* 30602 30701: gap of 100 bp
* 30702 31467: contig of 766 bp in length
* 31468 31567: gap of 100 bp
* 31568 32343: contig of 776 bp in length
* 32344 32443: gap of 100 bp
* 32444 33233: contig of 790 bp in length
* 33234 33333: gap of 100 bp
* 33334 34108: contig of 775 bp in length
* 34109 34208: gap of 100 bp
* 34209 34982: contig of 774 bp in length
* 34983 35082: gap of 100 bp
* 35083 35863: contig of 781 bp in length
* 35864 36716: contig of 753 bp in length
* 36717 36816: gap of 100 bp
* 36817 37591: contig of 775 bp in length
* 37592 37691: gap of 100 bp
* 37692 38467: contig of 776 bp in length
* 38468 38567: gap of 100 bp
* 38568 39336: contig of 769 bp in length
* 39337 39436: gap of 100 bp
* 39437 40184: contig of 748 bp in length
* 40185 40284: gap of 100 bp
* 40285 41091: contig of 807 bp in length
* 41092 41191: gap of 100 bp
* 41192 41942: contig of 751 bp in length
* 41943 42042: gap of 100 bp
* 42043 42823: contig of 781 bp in length
* 42824 42923: gap of 100 bp
* 42924 43708: contig of 785 bp in length
* 43709 43808: gap of 100 bp
* 43809 44580: contig of 772 bp in length
* 44581 44680: gap of 100 bp
* 44681 45425: contig of 745 bp in length
* 45426 45525: gap of 100 bp
* 45526 46309: contig of 784 bp in length
* 46310 46409: gap of 100 bp
* 46410 47193: contig of 784 bp in length
* 47194 47293: gap of 100 bp
* 47294 48062: contig of 773 bp in length
* 48067 48166: gap of 100 bp
* 48167 48931: contig of 765 bp in length
* 48932 49031: gap of 100 bp
* 49032 49812: contig of 781 bp in length
* 49813 49912: gap of 100 bp
* 49913 50664: contig of 752 bp in length
* 50665 50764: gap of 100 bp
* 50765 51522: contig of 758 bp in length
* 51523 51622: gap of 100 bp
* 51623 52382: contig of 760 bp in length
* 52383 52482: gap of 100 bp
* 52483 53264: contig of 782 bp in length

* 53265 53364: gap of 100 bp
 * 53365 54144: contig of 780 bp in length
 * 54145 54244: gap of 100 bp
 * 54245 55007: contig of 763 bp in length
 * 55008 55107: gap of 100 bp
 * 55108 55883: contig of 776 bp in length
 * 55884 55983: gap of 100 bp
 * 55984 56757: contig of 774 bp in length
 * 56758 56857: gap of 100 bp
 * 56858 57632: contig of 775 bp in length
 * 57633 57732: gap of 100 bp
 * 57733 58512: contig of 780 bp in length
 * 58513 58612: gap of 100 bp
 * 58613 59370: contig of 758 bp in length
 * 59371 59470: gap of 100 bp
 * 59471 60240: contig of 770 bp in length
 * 60241 60340: gap of 100 bp

Query Match 3.8%; Score 41.8; DB 2; Length 76141;
 Best Local Similarity 46.3%; Pred. No. 13;
 Matches 100; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 596 ACAATCCTGAAACGGGATAATAGAAATAATGTTCTTAATTCGAAGAAATTTAATGACT 655
 DB 40996 AAGATATTTAAATCTTATAAAGAAATAATAATAATTTATAATAATTTAATAAT 40937

QY 656 GCTTGTGCTTAAGAAATCAGAAAGAGACCGAAAGCTTCTAAGTAATCAGAAAAAT 715
 DB 40936 TATTAATAAAAAAANNAAAAAATTAATAAATNNNNNTTAATAAANNAGAAA 40877

QY 716 CTTCGACAAATAGCTTTCCATAGACAAACCGTGCAGAAAGGTCGATACCAATAATTC 775
 DB 40876 ATANTAGATNTNNNTGGGTTTAAAAAATAAAGAAATAAAAAAATAAAAAAAT 40817

QY 776 CGATTAATAAAGGAGTGCTTACACACCCATATCA 811
 DB 40816 NAATAAATAAAAAATTAATAAAAAAATAAAAAAATA 40781

RESULT 39
 AC105744
 LOCUS
 DEFINITION
 AC105744
 AC105744.7 GI:20303637
 HTG; HTGS_PHASE2.
 SOURCE
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 115666)
 Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Tallon, L., Gansberger, K.,
 Kim, M., Overton, L.L., Bera, J., Tsitrin, T., Krol, M., Jarrahi, B.,
 Jin, S., Koo, H., Zismann, V., Hsiang, J., Blunt, S., Vanaken, S.,
 Uterbach, T., Feldblyum, T., Yang, Q., Haas, B., Suh, B., Peterson, J.,
 Quackenbush, J., White, O., Salzberg, S. and Fraser, C.
 Oryza sativa ssp. japonica cv. Nipponbare OSJNB0009C07 BAC genomic
 sequence
 Unpublished
 2 (bases 1 to 115666)
 Buell, R.
 Direct Submission
 Submitted (09-JAN-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 3 (bases 1 to 115666)
 Buell, R.
 Direct Submission
 Submitted (25-APR-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 JOURNAL
 COMMENT
 On Apr 25, 2002 this sequence version replaced gi:20153307.
 * NOTE: This is a working draft sequence. It currently

* consists of 8 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 18477: contig of 18477 bp in length
 * 18478 18578: gap of unknown length
 * 18578 20885: contig of 2308 bp in length
 * 20886 20986: gap of unknown length
 * 20986 25966: contig of 4981 bp in length
 * 25967 26067: gap of unknown length
 * 26067 49469: contig of 23403 bp in length
 * 49470 49569: gap of unknown length
 * 49570 52659: contig of 3090 bp in length
 * 52660 82661: gap of unknown length
 * 82661 82961: contig of 30102 bp in length
 * 82962 110436: contig of 27475 bp in length
 * 110437 110536: gap of unknown length
 * 110537 115666: contig of 5130 bp in length.

FEATURES

source
 1. 115666
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="10"
 /clone="OSJNB0009C07"
 /note="Japonica cultivar-group"
 BASE COUNT 32490 a 24957 c 25425 g 31751 t 1043 others
 ORIGIN

Query Match 3.8%; Score 41.8; DB 2; Length 115666;
 Best Local Similarity 12.3%; Pred. No. 12;
 Matches 37; Conservative 136; Mismatches 128; Indels 0; Gaps 0;

QY 442 AGAAATGAGGGATGTAGAACCAAGGACGTCGCCCTAAAAGTCATGATGAGATCAG 501
 DB 46372 RKGKMGKSGSGSSGCGYKMGSGRKGSRKMSRRGRTKSGMSYKWSWSKS 46431

QY 502 AGTTAAGAGAAATATCTTATCGAGAGTTCGCATTACAGAGCAATGGAAAGTAACG 561
 DB 46432 RKGGMNTSYWMMYTCRRSMYMSRMMYMMYMKMSYKRYSKAMKSKAGWSK 46491

QY 562 GTCCTGCAACAGACAGCAAGTATTGATGGGGAGCTACAATCTCGAAACGGATAATAGGA 621
 DB 46492 KRCKSKWKAKSRMVKMTYWKWSSSCHWRTWYWMKCCMKWKMTAWKWM 46551

QY 622 AAATGTGTTTCTAATTCGAGAAATTTAATGACTCTTGATGCTTAAGAAATCAGAAAG 581
 DB 46552 WMRWKKSWKATARRMWMWRCWGWRRRRSAGRRRCSSNSMKRKSWSKS 46611

QY 682 GAGACCGAAAGCTTTCTAAGTAATCAAGAAAAAATCTTGCAACAATAGCTTTTCCATAGAGA 741
 DB 46612 SACMCMYMYACACSRWYMMKMAAAGARWKMAGAGAGWAGTGTATGTAGTAGTRA 46671

QY 742 A 742
 DB 46672 A 46672

RESULT 40

AL391883
 LOCUS
 DEFINITION
 AC391883
 AC391883.16 GI:16972968
 HTG.
 SOURCE
 ORGANISM
 Human DNA sequence from clone RP11-91K11 on chromosome 1, complete
 sequence.
 Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. Ellington, A. Direct Submission

Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequest@sanger.ac.uk humquery@sanger.ac.uk This sequence version replaced gi:15131273. On Nov 16, 2001 this sequence version replaced gi:15131273. During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at: http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGB/Chrl> RP11-91K11 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-91K11 It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-91K11 is at 126371 in this sequence. The true left end of clone RP11-460G22 is at 31870 in this sequence. The true right end of clone RP1-301K23 is at 2000 in this sequence.

FEATURES
source
1..126371
/location=Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-91K11"
/clone_lib="RP11-11.1"

BASE COUNT 35354 a 27351 c 28154 g 35512 t

ORIGIN

Query Match 3.8%; Score 41.8; DB 9; Length 126371;
Best Local Similarity 47.8%; Pred. No. 12;
Matches 121; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 698 AGTAATCAAGAAAAATCTTGGACATAGCTTTCCATAGAGAAACCGTGAGGAAGG 757
Db 45012 AAGGAATAATAACAATAATTTGTATATAGGTATGACACATCTAAGAATTTBAGG 45071

QY 758 AGTCGATACCAATAATCCGATTAAAAAGGGAGTGCTTTACCAACAGCCATVACGTCGAG 817
Db 45072 AGAAGTGATTGGTATCTGCAATTTACTTTGAATTACATAAAAAAATAAGATGAGTGTAT 45131

QY 818 CAACCTTCCTCCTTTGATCTTTGCATAATGACAGAGATGACAAAGAGATGAGTCTTTTGA 877
Db 45132 GAATGGTTAGAAGGATGATTAGATGAACCAATGAAGCAAGTATATAGATATTGATGCTA 45191

QY 878 AATCTCTAAATAGAAATGCATTTGTGTCGAGCTTAAAGCTTGTCTCTCTTTATTTCTCTTT 937
Db 45192 GAATCTAGACGGTGGATATATTGGTATGTACCATAACTTCTCTCTCTCTCTCTCTTTT 45251

QY 938 TGTAGATGATTC 950

Db 45252 TACACAGTCTTGC 45264

RESULT 41
AC004153 156060 bp DNA linear HTG 12-AUG-2000
LOCUS Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 unordered pieces.
AC004153
AC004153.7 GI:97977733
AC004153.7 GI:97977733
HTG; HTGS_PHASE1.
SOURCE Plasmodium falciparum.
ORGANISM Plasmodium falciparum.
REFERENCE
1 (bases 1 to 156060)
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
Kurd,I.O.B., Conway,A.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 156060)
REFERENCE Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
AUTHORS Submitted (18-FEB-1998) Stanford DNA Sequencing and Technology
JOURNAL Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA

COMMENT
On Aug 12, 2000 this sequence version replaced gi:8810454.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 102169: contig of 102169 bp in length
* 102170 102369: gap of unknown length
* 102370 156060: contig of 53691 bp in length.

FEATURES
source
1..156060
/location=Qualifiers
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PFYAC812"
/clone="3D7"

BASE COUNT 62615 a 14889 c 15137 g 63219 t 200 others

ORIGIN

Query Match 3.8%; Score 41.8; DB 2; Length 156060;
Best Local Similarity 45.3%; Pred. No. 12;
Matches 151; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 568 CAACAGACACAAGTATTGATGGGAGCTACAATCTGAAACGGGATTAATAGGAAATG 627
Db 110223 CAAAAAATAAATAATGATGATTAATTCAGATATGGAATGGAAAAATTTCAATCCAC 110282

QY 628 TGTCTTAAATCGAAGAAATTTAATGACTGCTTGATGCTTAAGAATTCAGAAAGGAGACC 687
Db 110283 AATTACATATTTCAATACCTTTTGNACCTAATGAAGATATGTTCTTTCTAAAGAAAAAC 110342

QY 688 GAAGCTTCTTAAGTAATCAAGAAAAAATCTTGGACATAAGCTTTTCCATAGAGAAAAACCG 747
Db 110343 ATGATATCTCAATGATGGAGAAAAAAGGGTCTTAATATATGCAAAAAAACAATCCA 110402

QY 748 TGAGGAAGGAGTCGATACCATTAATTCGATTTAAAAAGGGAGTGCTTTACAACAGCAT 807
Db 110403 AAAGAAAAAGTCTAGATTAATCAAAATGATTAATAGAAATATAAATAAATACATTA 110462

QY 808 ATCAGTGGAGCAACCTTCCCTTTTGATTTGATTTGATTAATGACAGAAGTGAACAAGATAA 867
Db 110463 ATAATAATGAAGAAAAAGATCATGCTTTTATCAATGAAATTCATTAATAAAAA 110522

QY 868 GTTTTGAATAATCTTCAAAATAGAAATGCATTT 900

Db 13383 TACACACTCTTGC 13395

RESULT 44
AC109165/c
LOCUS
DEFINITION
AC109165
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
TITLE
JOURNAL
REFERENCE
AUTHORS

200279 bp DNA linear HTG 25-APR-2002
Mus musculus clone RP23-30B3, WORKING DRAFT SEQUENCE, 16 ordered
pieces.
AC109165
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 200279)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-30B3
Unpublished
2 (bases 1 to 200279)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemke, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 200279)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepe, Y., Colangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zemke, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 25, 2002 this sequence version replaced gi:20258472.

FEATURES
source
1..200279
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-30B3"
/clone_lib="RP23-30B3 Female Mouse BAC"
misc_feature
1..564
misc_feature
665..1284
/note="assembly_fragment"
/note="assembly_fragment"

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L18963
Center clone name: 30_B_3
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 19597 bases at least Q40
Consensus quality: 197569 bases at least Q30
Consensus quality: 198340 bases at least Q20
Insert size: 198000; agarose-fp
Quality coverage: 7.0 in Q20 bases; agarose-fp
Quality coverage: 7.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 564: contig of 564 bp in length
* 665 1284: contig of 620 bp in length
* 1285 1384: gap of 100 bp
* 1385 2041: contig of 657 bp in length
* 2042 2141: gap of 100 bp
* 2142 3109: contig of 968 bp in length
* 3110 3209: gap of 100 bp
* 3210 4516: contig of 1307 bp in length
* 4517 4616: gap of 100 bp
* 4617 6565: contig of 1949 bp in length
* 6566 6665: gap of 100 bp
* 6666 9499: contig of 2834 bp in length
* 9500 9599: gap of 100 bp
* 9600 13090: contig of 3491 bp in length
* 13091 13190: gap of 100 bp
* 13191 17101: contig of 3911 bp in length
* 17202 22059: contig of 4858 bp in length
* 22060 22159: gap of 100 bp
* 22160 30887: contig of 8728 bp in length
* 30888 30887: gap of 100 bp
* 30888 42016: contig of 11029 bp in length
* 42017 42116: gap of 100 bp
* 42117 54621: contig of 12505 bp in length
* 54622 54721: gap of 100 bp
* 54722 80967: contig of 26246 bp in length
* 80968 81067: gap of 100 bp
* 81068 136318: contig of 55251 bp in length
* 136319 136418: gap of 100 bp
* 136419 200279: contig of 63861 bp in length.
Location/Qualifiers
1..200279
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-30B3"
/clone_lib="RP23-30B3 Female Mouse BAC"
misc_feature
1..564
misc_feature
665..1284
/note="assembly_fragment"
/note="assembly_fragment"

TITLE	Direct Submission	22154	contig of 1061 bp in length
JOURNAL	Unpublished	23215	gap of unknown length
REFERENCE	2 (bases 1 to 103337)	23315	contig of 1570 bp in length
AUTHORS	Worley, K.C.	24884	gap of unknown length
TITLE	Direct Submission	24985	contig of 1649 bp in length
JOURNAL	Submitted (19-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	26733	gap of unknown length
REFERENCE	3 (bases 1 to 103337)	26734	contig of 1831 bp in length
AUTHORS	Worley, K.C.	28564	gap of unknown length
TITLE	Direct Submission	28565	contig of 1903 bp in length
JOURNAL	Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	30567	gap of unknown length
COMMENT	On Jul 12, 2002 this sequence version replaced gi:19549112. ----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GNZA Center clone name: CH230-74P14 ----- Summary Statistics Sequencing vector: Plasmid Chemistry: Dye-terminator Big Dye: 100% of reads Assembly program: Phrap: version 0.990329 Consensus quality: 50058 bases at least Q40 Consensus quality: 53149 bases at least Q30 Consensus quality: 56140 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 50 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * as soon as it is available and the accession number will * be preserved. * * 1068: contig of 1068 bp in length * 1069 * 1169: gap of unknown length * 2334 * 2433: contig of 1165 bp in length * 2434 * 3443: gap of unknown length * 3444 * 3543: contig of 1010 bp in length * 3544 * 4553: gap of unknown length * 4556 * 4635: gap of unknown length * 6147: contig of 1012 bp in length * 6148 * 6247: gap of unknown length * 6248 * 7409: contig of 1162 bp in length * 7410 * 7509: gap of unknown length * 7510 * 9245: contig of 1736 bp in length * 9246 * 9345: gap of unknown length * 9346 * 10347: contig of 1002 bp in length * 10348 * 10447: gap of unknown length * 10448 * 11789: contig of 1342 bp in length * 11790 * 11889: gap of unknown length * 11890 * 12964: contig of 1075 bp in length * 12965 * 13064: gap of unknown length * 13065 * 14595: contig of 1531 bp in length * 14596 * 14695: gap of unknown length * 14696 * 16163: contig of 1468 bp in length * 16164 * 16263: gap of unknown length * 16264 * 17419: contig of 1156 bp in length * 17420 * 17519: gap of unknown length * 17520 * 19034: contig of 1515 bp in length * 19035 * 19134: gap of unknown length * 19135 * 20254: contig of 1120 bp in length * 20255 * 20354: gap of unknown length * 20355 * 22053: contig of 1699 bp in length * 22054 * 22153: gap of unknown length		

ORIGIN

```
Query Match      3.8%; Score 41.6; DB 2; Length 103337;
Best Local Similarity 51.6%; Pred. No. 14;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 678 AAGGACCGAAGGCTTCTAAGTAATCAAGAAAAATCTTGGACAAATAGCTTTTCCATA 737
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 57477 AAGAAAAGCTGCTTTTGGAGAGATTAAACAAGATAGATAAAACCCCTTAGCAAAATCTATC 57536
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 738 GAGAAAACCGTGAGGAAAGGAGTCATACCAATTAATCCGATTAAAAAAGGAGTGCTTTA 797
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 57537 TAAAGGCTCCAGAGGAAATTTCCCAAATTCAGAAAATCAGATATGAARAGGTAGAGTAG 57596
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 798 CAACAGCATATCAGTGGAGCAACCTTCCCTTGTATCTTTGATATGACAGAGATGAA 857
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 57597 CACTGATATAGGAGGATTCAAAATAATCATCAGATCTTACTACAAAAGCTTATACTGAA 57656
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 858 CAAA 861
Db 57657 CAAA 57660

RESULT 48
AL353694_1
WPCOMMENT
Sequence split into 4 fragments LOCUS AL353694 Accession AL353694
Fragment Name      Begin      End
AL353694.0         1      110000
AL353694.1        100001      210000
AL353694.2        200001      310000
AL353694.3        300001      356330
Continuation (2 of 4) of AL353694 from base 100001 (AL353694 Homo sapiens chromosome 6 c
Query Match      3.8%; Score 41.6; DB 2; Length 110000;
Best Local Similarity 54.6%; Pred. No. 14;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 659 AGAATCAGAAAGGACCGAAGCTTCTAAGTAATCAAGAAAAATCTTGGACAAATAGC 728
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 91714 ACAAACATCAATGAACAAAAAATTTGTTTTTTTGAAGATATAAATTTGGACAAACCA 91773
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 729 TTTTCCATAGAGAAACCGTGAGGAAAGGAGTCGATACCAATAATTCGATTAAAAAAGG 788
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 91774 TTAGCCAGACAAAAAGAGAGAGAGAGAACCCCAATAAACAAAAACAGAGATAAAAAAAG 91833
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 789 GAGTGCCTACACAGCCATATCAGTGGAGCAA 820
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 91834 GAGACATTACAACTGATACCCACAGAAATTCAA 91865
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 49
AL354818/c
LOCUS
DEFINITION
Homo sapiens chromosome 13 clone RP11-78L16, ** SEQUENCING IN
PROGRESS ***, 25 unordered pieces.
ACCESSION
AL354818
VERSION
AL354818.3 GI:10039795
KEYWORDS
HTG; HTGS_PHASE1; HTGS-CANCELLED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148549)
Direct Submission
AUTHORS
Burton,J.
TITLE
Submitted (10-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
JOURNAL
requests: clonerequest@sanger.ac.uk
COMMENT
On Sep 9, 2000 this sequence version replaced gi:9581627.
----- Genome Center
Center: Sanger Centre
Center code: SC
```

Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA78L16
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 13724 bases at least Q40
Consensus quality: 139166 bases at least Q30
Consensus quality: 142544 bases at least Q20
Insert size: 146149; sum-of-contigs
Insert size: 153445; 5.9% error; agarose-fp
Quality coverage: 2.84x in Q20 bases; sum-of-contigs Quality
coverage: 3.00x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 4894: contig of 4894 bp in length
4895 4994: gap of 100 bp
4995 16378: contig of 11384 bp in length
16379 16478: gap of 100 bp
16479 18490: contig of 2012 bp in length
18491 18590: gap of 100 bp
18591 21846: contig of 3256 bp in length
21847 21946: gap of 100 bp
21947 26998: contig of 5052 bp in length
26999 27098: gap of 100 bp
27099 29374: contig of 2276 bp in length
29375 29474: gap of 100 bp
29475 32601: contig of 3127 bp in length
32602 32701: gap of 100 bp
32702 38362: contig of 5661 bp in length
38363 38462: gap of 100 bp
38463 48272: contig of 9810 bp in length
48273 48372: gap of 100 bp
48373 52389: contig of 4017 bp in length
52390 52489: gap of 100 bp
52490 62396: contig of 9907 bp in length
62397 62496: gap of 100 bp
62497 67828: contig of 5332 bp in length
67829 67928: gap of 100 bp
67929 70816: contig of 2891 bp in length
70820 70919: gap of 100 bp
70920 79571: contig of 8652 bp in length
79572 79671: gap of 100 bp
79672 83474: contig of 3803 bp in length
83475 83574: gap of 100 bp
83575 90030: contig of 6456 bp in length
90031 90130: gap of 100 bp
90131 112698: contig of 22568 bp in length
112699 112798: gap of 100 bp
112799 118187: contig of 5389 bp in length
118188 118287: gap of 100 bp
118288 122327: contig of 4040 bp in length
122328 122427: gap of 100 bp
122428 130709: contig of 8282 bp in length
130710 130809: gap of 100 bp
130810 132827: contig of 2018 bp in length
132828 132927: gap of 100 bp
132928 135248: contig of 3321 bp in length
135249 136348: gap of 100 bp
136349 139531: contig of 3183 bp in length
139532 139631: gap of 100 bp
139632 143470: contig of 3839 bp in length
143471 143570: gap of 100 bp
143571 148549: contig of 4979 bp in length.
```


FEATURES
source

Location/Qualifiers
 1. .148549
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="13"
 /clone="RP11-78L16"
 /clone_lib="RPC1-11.1"
 1. .4894
 /note="assembly_fragment:01091
 fragment_chain:1"
 4995. .16378
 /note="assembly_fragment:00072
 fragment_chain:1"
 16479. .18490
 /note="assembly_fragment:00411
 fragment_chain:1"
 18591. .21846
 /note="assembly_fragment:01041
 fragment_chain:1"
 21947. .26998
 /note="assembly_fragment:00977
 fragment_chain:1"
 27099. .29374
 /note="assembly_fragment:00091
 fragment_chain:2"
 29475. .32601
 /note="assembly_fragment:00282
 fragment_chain:2"
 32702. .38362
 /note="assembly_fragment:00258
 fragment_chain:3"
 38463. .48272
 /note="assembly_fragment:00607
 fragment_chain:3"
 48373. .52389
 /note="assembly_fragment:00980
 fragment_chain:4"
 52490. .62396
 /note="assembly_fragment:00644
 fragment_chain:4"
 62497. .67828
 /note="assembly_fragment:01219
 fragment_chain:5"
 67929. .70819
 /note="assembly_fragment:00679
 fragment_chain:5"
 70320. .79571
 /note="assembly_fragment:00075"
 79672. .83474
 /note="assembly_fragment:00377"
 83575. .90030
 /note="assembly_fragment:00645"
 90131. .112598
 /note="assembly_fragment:00685"
 112799. .118187
 /note="assembly_fragment:00757"
 118288. .122327
 /note="assembly_fragment:00791"
 122428. .130709
 /note="assembly_fragment:00937"
 130810. .132827
 /note="assembly_fragment:01022"
 132928. .136248
 /note="assembly_fragment:01160"
 136349. .139531
 /note="assembly_fragment:01267"
 139632. .143470
 /note="assembly_fragment:00779
 fragment_chain:6"
 143571. .148549
 /note="assembly_fragment:00210
 fragment_chain:6
 clone_end:T

BASE COUNT 43637 a 27779 c 27991 g 46687 t 2455 others
 ORIGIN
 Query Match 3.8%; Score 41.6; DB 2; Length 148549;
 Best Local Similarity 48.6%; Pred. No. 13; Gaps 1;
 Matches 142; Conservative 0; Mismatches 149; Indels 1;
 QY 598 AATCTCTGAACGGGTAATAGGAAAATGTGTTTCTAATTCGAGAATTTAATGACTGC 657
 DB 31084 AAAGATTAACTGGGATATTCAGATTAATAAGATAAAAAACCTTATGCTATTTCATAGA 31025
 QY 658 TTGATGCTCTAGAAAATCAGAAAGGAGACCGAAGCTCTTAAGTAATCAAGAAAAATCT 717
 DB 31024 AGGATGAATGGAACACACAGACATTTATTTTAAAGAGTATTAAGGCTCTTATACATATGT 30965
 QY 718 TGGCAATAGCTTTTTCATAGAGAAAACCGTCAGGAAGGAGTCGATACCAATAATTCGG 777
 DB 30964 TACAATATCAAAATTTATCTATATATATCCAGTGTGTAAGTCTCTAAGATTAATTCCT 30905
 QY 778 ATTAAGAAAGGGAGTGTCTTACACAGCCATATCAGTGGAGCAACCTTCCCTTTGATTCCT 837
 DB 30904 GCTCTACTGTATAAATATATAAAAACITTAATTAGGACAAAGCATCTTTTAGAGGCTA 30845
 QY 838 T-GATAATGACAGAAGATGAACAAAGAGTAAAGTTTGTGAAAAATTTCTAAAT 888
 DB 30844 GAGATAAAACCACACTAAACAAACAGGAAACCTTTAAGTTTCAACAT 30793
 RESULT 50
 AC046180 215896 bp DNA linear HTG 12-MAY-2000
 AC046180 Homo sapiens chromosome 15 clone RP11-465J17 map 15, WORKING DRAFT
 DEFINITION
 AC046180
 AC046180
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 Homo sapiens
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 215896)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 15, clone RP11-465J17
 2 (bases 1 to 215896)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boockvar,L., Bouckhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardina,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Lakocque,K., Lamazares,R., Landers,T., Lehotsky,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
 Meidrum,J., Meneus,L., Mihoval,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triglio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 12, 2000 this sequence version replaced gi:7547220.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8718
Center clone name: 465_J_17
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 203149 bases at least Q40
Consensus quality: 209316 bases at least Q30
Consensus quality: 211902 bases at least Q20
Insert size: 220000; agarose-fp
Insert size: 213296; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1011: contig of 1011 bp in length
* 1012 1111: gap of 100 bp
* 1112 2633: contig of 1522 bp in length
* 2634 2733: gap of 100 bp
* 2734 4093: contig of 1360 bp in length
* 4094 4193: gap of 100 bp
* 4194 5298: contig of 1105 bp in length
* 5299 5398: gap of 100 bp
* 5399 7088: contig of 1690 bp in length
* 7089 7188: gap of 100 bp
* 7189 7889: contig of 701 bp in length
* 7890 7989: gap of 100 bp
* 7990 10304: contig of 2315 bp in length
* 10305 10404: gap of 100 bp
* 10405 12955: contig of 2551 bp in length
* 12956 13053: gap of 100 bp
* 13056 16426: contig of 3371 bp in length
* 16427 16526: gap of 100 bp
* 16527 19178: contig of 2652 bp in length
* 19179 19278: gap of 100 bp
* 19279 23519: contig of 4241 bp in length
* 23520 26853: contig of 3234 bp in length
* 26854 26954: gap of 100 bp
* 26954 31292: contig of 4339 bp in length
* 31293 31392: gap of 100 bp
* 31393 36401: contig of 5009 bp in length
* 36402 36501: gap of 100 bp
* 36502 41321: contig of 4820 bp in length
* 41322 41421: gap of 100 bp
* 41422 47286: contig of 5865 bp in length
* 47287 47386: gap of 100 bp
* 47387 54497: contig of 7111 bp in length
* 54498 54597: gap of 100 bp
* 54598 60757: contig of 6160 bp in length
* 60758 60857: gap of 100 bp
* 60858 68004: contig of 7147 bp in length
* 68005 68104: gap of 100 bp
* 68105 77167: contig of 9063 bp in length
* 77168 77267: gap of 100 bp
* 77268 89567: contig of 11300 bp in length
* 89568 88667: gap of 100 bp
* 88668 101419: contig of 12752 bp in length
* 101420 101519: gap of 100 bp

```

```

* 101520 114909: contig of 13390 bp in length
* 114910 115009: gap of 100 bp
* 115010 130182: contig of 15173 bp in length
* 130183 130282: gap of 100 bp
* 130283 151503: contig of 21221 bp in length
* 151504 151603: gap of 100 bp
* 151604 179979: contig of 28376 bp in length
* 179980 180079: gap of 100 bp
* 180080 215896: contig of 35817 bp in length.
FEATURES
    Location/Qualifiers
        1..215896
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="15"
            /map="15"
            /clone="RP11-465J17"
            /clone_lib="RPC1-11 Human Male BAC"
        1..1011
            /note="assembly_fragment"
        1112..2633
            /note="assembly_fragment"
        2734..4093
            /note="assembly_fragment"
        4194..5298
            /note="assembly_fragment"
        5399..7088
            /note="assembly_fragment"
        7189..7889
            /note="assembly_fragment"
            clone_end:17
            vector_side:right
        7990..10304
            /note="assembly_fragment"
        10405..12955
            /note="assembly_fragment"
        13056..16426
            /note="assembly_fragment"
        16527..19178
            /note="assembly_fragment"
        19279..23519
            /note="assembly_fragment"
        23620..26853
            /note="assembly_fragment"
        26954..31292
            /note="assembly_fragment"
        31393..36401
            /note="assembly_fragment"
        36502..41321
            /note="assembly_fragment"
        41422..47286
            /note="assembly_fragment"
        47387..54497
            /note="assembly_fragment"
        54598..60757
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left
        60858..68004
            /note="assembly_fragment"
        68105..77167
            /note="assembly_fragment"
        77268..88567
            /note="assembly_fragment"
        88568..101419
            /note="assembly_fragment"
        101520..114909
            /note="assembly_fragment"
        115010..130182
            /note="assembly_fragment"
        130283..151503
            /note="assembly_fragment"
        151604..179979
            /note="assembly_fragment"

```

us-09-438-185a-1_1199590_1200675.rge

Fri Jan 10 12:01:26 2003

```
misc_feature 180080. 215896
BASE COUNT 66799 a 38601 c 40541 g 67354 t 2601 others
ORIGIN
Query Match 3.8%; Score 41.6; DB 2; Length 215896;
Best Local Similarity 48.6%; Pred. No. 13;
Matches 142; Conservative 0; Mismatches 149; Indels 1; Gaps 1;
QY 598 AAATCCTGAAACGGGATAATAGGAAAAATGTGTTCTTAATTCGAGAAATTTAATGACTGC 657
Db 113768 AANGATTTAACTGGGATATTCAGATTATAAGATAAAAAACCTTATGCTATTTTCATAGA 113827
QY 658 TTGATGGTCTAAGAAATCAGAAAGGAGACCGAAAGCTTCTAAGTAATCAAGAAAAAATCT 717
Db 113828 AGGATGAATGGAACACACAGACATTATTTAAGAGTATTAAAGGCCTTTATACATAIGTT 113887
QY 718 TGGACAATAGCTTTTCCATAGAGAAAAACCGTGAGGAAAGGAGTCGATACATAAATTCGG 777
Db 113888 TACATATCAAAATTTATCTATATATATCCAGTGTGAACTGAGTCTCTAAGATTAAATTCCT 113947
QY 778 AATTAAAAAGGCGACTGCTTACACACGCCATATCAGTGGAGCAACCTTCCTCTTGATTCCT 837
Db 113948 GCTCTACTACTGTATATAATATATAAAACCTTTAATTAGRCNAGCATCTTTTAGAGGCTA 114007
QY 838 T-GATAATGACAGAGATGACACAGAGTAAGTTTTTGAAAAATTCFAAAT 888
Db 114008 GAGATAAAACCACACTAAAAAACCAAGGAAAAACCTTTAAAGTTTCAACAT 114059
```

Search completed: January 9, 2003, 12:09:21
Job time : 6395 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 16:59:59 ; Search time 41 Seconds

(without alignments)
171.297 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKALKL.....ESIPLYNQEKYLSGFVLCQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216.5	11.5	484	9 US-10-154-674-8	Sequence 8, Appli
2	216.5	11.5	485	9 US-10-154-674-6	Sequence 6, Appli
3	216.5	11.5	486	9 US-10-154-674-4	Sequence 4, Appli
4	216.5	11.5	490	9 US-10-154-674-2	Sequence 2, Appli
5	206.5	10.9	532	10 US-09-205-658-313	Sequence 313, App
6	95	5.0	1427	10 US-09-801-368-354	Sequence 354, App
7	92	4.9	1427	10 US-09-817-774-23	Sequence 23, Appl
8	92	4.9	4563	9 US-09-870-759-128	Sequence 128, App
9	88.5	4.7	511	10 US-09-746-491-37	Sequence 37, Appl
10	88.5	4.7	550	10 US-09-746-491-40	Sequence 40, Appl
11	88.5	4.7	541	10 US-09-746-491-39	Sequence 39, Appl
12	87.5	4.6	541	10 US-09-881-421-4	Sequence 4, Appli
13	87.5	4.6	541	12 US-10-157-447-2	Sequence 2, Appli
14	84	4.4	978	10 US-09-864-761-43207	Sequence 43207, A
15	83.5	4.4	450	10 US-09-741-669-385	Sequence 385, App
16	83.5	4.4	835	10 US-08-801-574-23	Sequence 22, Appl
17	82.5	4.4	995	10 US-09-864-761-45017	Sequence 45017, A
18	82.5	4.4	1202	10 US-09-864-761-43061	Sequence 43061, A
19	82	4.3	909	10 US-09-815-242-12990	Sequence 12990, A

4.3	818	10	US-09-833-790-366	Sequence 366, App
4.3	1216	10	US-09-815-242-13312	Sequence 13312, A
4.3	1216	10	US-09-815-242-13698	Sequence 13698, A
4.3	3056	8	US-08-984-090-2	Sequence 2, Appli
4.3	323	10	US-09-801-368-352	Sequence 352, App
4.3	474	10	US-09-765-272-60	Sequence 60, Appl
4.3	823	9	US-09-895-913A-244	Sequence 244, App
4.3	1993	9	US-10-098-979-2	Sequence 2, Appli
4.3	197	10	US-09-764-853-836	Sequence 836, App
4.3	736	9	US-09-978-295A-526	Sequence 526, App
4.3	736	9	US-09-978-697-526	Sequence 526, App
4.3	736	9	US-09-978-192A-526	Sequence 526, App
4.3	736	9	US-09-999-832A-526	Sequence 526, App
4.3	736	9	US-09-978-189-526	Sequence 526, App
4.3	736	12	US-10-052-586-420	Sequence 420, App
4.3	765	10	US-09-888-615-80	Sequence 80, Appl
4.2	560	10	US-09-815-242-13304	Sequence 13304, A
4.2	897	10	US-09-815-242-11609	Sequence 11609, A
4.2	645	10	US-09-731-231A-2	Sequence 2, Appli
4.2	867	9	US-09-932-598-84	Sequence 84, Appl
4.2	867	9	US-09-989-293A-84	Sequence 84, Appl
4.2	867	9	US-09-989-735-84	Sequence 84, Appl
4.2	867	9	US-09-990-444-84	Sequence 84, Appl
4.2	867	9	US-09-989-730-84	Sequence 84, Appl
4.2	867	9	US-09-990-436-84	Sequence 84, Appl
4.2	867	9	US-09-991-181-84	Sequence 84, Appl
4.2	867	9	US-09-993-687-84	Sequence 84, Appl
4.2	867	9	US-09-989-734-84	Sequence 84, Appl
4.2	867	9	US-10-028-072-20	Sequence 20, Appl
4.2	867	10	US-09-989-723-84	Sequence 84, Appl
4.2	867	10	US-09-989-723-84	Sequence 84, Appl
4.2	867	10	US-09-989-279-84	Sequence 84, Appl
4.2	867	10	US-09-989-731-84	Sequence 84, Appl
4.2	867	10	US-09-989-732-84	Sequence 84, Appl
4.2	867	10	US-09-991-073-84	Sequence 84, Appl
4.2	867	10	US-09-990-442-84	Sequence 84, Appl
4.2	867	10	US-09-991-163-84	Sequence 84, Appl
4.2	867	10	US-09-993-604-84	Sequence 84, Appl
4.2	867	10	US-09-990-456-84	Sequence 84, Appl
4.2	867	10	US-09-989-721-84	Sequence 84, Appl
4.2	867	10	US-09-881-752A-140	Sequence 140, App
4.2	410	10	US-09-881-752A-140	Sequence 140, App
4.1	185	9	US-10-007-270-15	Sequence 15, Appl
4.1	328	10	US-09-942-446-4	Sequence 4, Appli
4.1	328	10	US-09-995-693-35	Sequence 35, Appl
4.1	328	10	US-09-938-970-4	Sequence 4, Appli
4.1	506	10	US-09-815-242-5045	Sequence 5045, A
4.1	506	10	US-09-815-242-11757	Sequence 11757, A
4.1	790	10	US-09-801-368-184	Sequence 184, App
4.1	639	10	US-09-815-242-5390	Sequence 5390, Ap
4.1	646	10	US-09-815-242-12304	Sequence 12304, A
4.1	749	9	US-10-045-792-7	Sequence 7, Appli
4.1	749	10	US-09-833-205-6	Sequence 6, Appli
4.1	1663	10	US-09-875-519A-22	Sequence 22, Appl
4.1	1900	10	US-09-801-368-390	Sequence 390, App
4.1	311	9	US-10-027-805-41	Sequence 41, Appl
4.0	810	10	US-10-071-505-3	Sequence 3, Appli
4.0	745	10	US-09-815-242-12209	Sequence 12209, A
4.0	833	9	US-10-033-297-71	Sequence 71, Appl
4.0	342	9	US-09-847-010-4	Sequence 4, Appli
4.0	396	10	US-09-841-693-11	Sequence 11, Appl
4.0	407	10	US-09-841-693-9	Sequence 9, Appli
4.0	407	10	US-10-007-270-28	Sequence 28, Appl
4.0	777	9	US-10-007-270-2	Sequence 2, Appli
4.0	777	9	US-09-801-368-304	Sequence 304, App
4.0	1155	10	US-09-801-368-6	Sequence 6, Appli
4.0	1257	10	US-09-738-363-6	Sequence 118, App
4.0	1659	10	US-09-801-368B-118	Sequence 2, Appli
3.9	396	10	US-09-801-876B-2	Sequence 63, Appl
3.9	833	9	US-10-033-297-69	Sequence 56, Appl
3.9	1421	9	US-10-108-605-323	Sequence 323, App
3.9	2167	10	US-09-801-368-56	Sequence 56, Appl
3.9	263	10	US-09-864-761-33879	Sequence 33879, A

93 74 3.9 579 9 US-09-738-626-6373
94 74 3.9 579 10 US-09-865-825-2
95 74 3.9 579 10 US-09-865-825-5
96 73.5 3.9 229 10 US-09-828-644-113
97 73.5 3.9 321 10 US-09-815-242-10490
98 73.5 3.9 331 10 US-09-886-055-145
99 73.5 3.9 510 10 US-09-923-246-89
100 73.5 3.9 547 10 US-09-961-527A-11
101 73.5 3.9 560 9 US-10-071-505-2
102 73.5 3.9 832 9 US-10-033-297-4
103 73.5 3.9 832 9 US-09-972-834-2
104 73.5 3.9 832 9 US-10-071-505-1
105 73.5 3.9 832 9 US-10-081-806-4
106 73.5 3.9 833 9 US-10-033-297-66
107 73 3.9 167 9 US-10-001-857-192
108 73 3.9 585 10 US-09-815-242-11753
109 73 3.9 623 10 US-09-764-870-276
110 73 3.9 643 9 US-10-063-547-132
111 73 3.9 643 9 US-10-004-551-6
112 73 3.9 649 12 US-10-006-867-132
113 73 3.9 649 12 US-10-052-586-384
114 72.5 3.8 256 10 US-09-815-242-11326
115 72.5 3.8 394 10 US-09-925-302-722
116 72.5 3.8 560 10 US-09-801-574-53
117 72.5 3.8 569 12 US-10-078-929-116
118 72.5 3.8 639 10 US-09-789-417-2
119 72.5 3.8 639 10 US-09-792-451-2
120 72.5 3.8 639 12 US-10-052-586-364
121 72.5 3.8 708 10 US-09-735-705-369
122 72.5 3.8 708 10 US-09-850-716A-369
123 72.5 3.8 708 10 US-09-897-778-369
124 72.5 3.8 875 10 US-09-950-370-22
125 72.5 3.8 1167 10 US-09-815-242-11522
126 72 3.8 329 10 US-09-775-879-17
127 72 3.8 1648 10 US-09-515-806-4
128 71.5 3.8 457 9 US-09-738-626-4195
129 71.5 3.8 304 9 US-10-136-517-18
130 71 3.8 176 10 US-09-864-761-33919
131 71 3.8 309 10 US-09-886-055-489
132 71 3.8 677 10 US-09-745-763-168
133 71 3.8 759 10 US-09-864-864-893
134 71 3.8 268 10 US-09-841-132-432
135 70.5 3.7 321 10 US-09-886-055-117
136 70.5 3.7 351 10 US-09-935-943-26
137 70.5 3.7 354 10 US-09-729-674-174
138 70.5 3.7 524 10 US-09-925-300-1538
139 70.5 3.7 585 10 US-09-815-242-5426
140 70.5 3.7 585 10 US-09-815-242-12620
141 70.5 3.7 631 9 US-10-108-605-169
142 70.5 3.7 737 10 US-09-749-601A-9
143 70.5 3.7 884 9 US-09-843-676-55
144 70.5 3.7 884 9 US-09-766-253-55
145 70.5 3.7 1042 10 US-09-815-242-13345
146 70.5 3.7 1345 9 US-10-108-605-249
147 70 3.7 249 9 US-10-027-805-39
148 70 3.7 249 10 US-09-903-410-39
149 70 3.7 380 9 US-10-114-893-73
150 70 3.7 380 9 US-10-114-893-73

ALIGNMENTS

Sequence 6373, App
Sequence 2, Appl
Sequence 5, Appl
Sequence 113, Appl
Sequence 10490, A
Sequence 145, App
Sequence 89, Appl
Sequence 11, Appl
Sequence 2, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 4, Appl
Sequence 66, Appl
Sequence 152, App
Sequence 11753, A
Sequence 276, App
Sequence 132, App
Sequence 6, Appl
Sequence 132, App
Sequence 184, App
Sequence 1328, A
Sequence 722, App
Sequence 116, Appl
Sequence 2, Appl
Sequence 364, App
Sequence 369, App
Sequence 369, App
Sequence 369, App
Sequence 11522, A
Sequence 17, Appl
Sequence 4195, App
Sequence 18, Appl
Sequence 33919, A
Sequence 499, App
Sequence 168, App
Sequence 823, App
Sequence 117, App
Sequence 26, App
Sequence 174, App
Sequence 10, App
Sequence 1538, App
Sequence 5426, App
Sequence 12620, A
Sequence 169, App
Sequence 9, Appl
Sequence 55, Appl
Sequence 55, Appl
Sequence 13345, A
Sequence 249, App
Sequence 39, Appl
Sequence 39, Appl
Sequence 73, Appl

;; CURRENT APPLICATION NUMBER: US/10/154,674
;; CURRENT FILING DATE: 2002-05-23
;; PRIOR APPLICATION NUMBER: US 60/294,076
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 8
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 8
;; LENGTH: 484
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-154-674-8

Query Match 11.5%; Score 216.5; DB 9; Length 484;
Best Local Similarity 26.2%; Pred. No. 5.4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

117 LMKSYCRPFELDTLEAFGLSPDLH-----OAVIKPELETHTSYYPGVGVAAPHQ 168
221 LSKYPTTHACREYKFNPLITKTCYREDNVQLEDVSMFLKRSQGTVPVAGVLSPPD 280
169 YLSLQDRYPPIASVWKTLDKDNFSITPDLIHDLGHVPMILHPSSEFPINNRLFTKY 228
281 FLAGLAVRVFCTQYIRHGSDDLTPBPTCHELGHVPLADKXFAQSOEIG----- 334
229 IEKVQALPSKKQRIQTLSNLIAIVRCFWFVESGLIENHGRAYGAVLISSPOELGHA 288
335 ---LASGASDEVQKLT-----CYFTIEFGLCKQEGQLRAYGAGLSSIGBLKHA 384
289 FIDNRYVPLEDQIIRLPNTSTPOELTFSIRHDELVE-----L 329
385 LSDACVAPADPPTKTCLOECLITTFQEAVFSESEPEAKERMDFAKSITRPFVYFNBY 444
330 TSXLEMMDOGLESI 345
445 TQSEIILKDRSIEINV 460

US-10-154-674-6
;; Sequence 6, Application US/10154674
;; Publication No. US20020192694A1
;; GENERAL INFORMATION:
;; APPLICANT: Yu, Xuanchuan
;; APPLICANT: Miranda, Maricar
;; TITLE OF INVENTION: No. US20020192694A1 Human Hydroxylases and Polynucleotides
;; FILE REFERENCE: LEX-0352-USA
;; CURRENT APPLICATION NUMBER: US/10/154,674
;; CURRENT FILING DATE: 2002-05-23
;; PRIOR FILING DATE: 2001-05-29
;; NUMBER OF SEQ ID NOS: 8
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6
;; LENGTH: 485
;; TYPE: PRT
;; ORGANISM: homo sapiens
US-10-154-674-6

Query Match 11.5%; Score 216.5; DB 9; Length 485;
Best Local Similarity 26.2%; Pred. No. 5.4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;
117 LMKSYCRPFELDTLEAFGLSPDLH-----OAVIKPELETHTSYYPGVGVAAPHQ 168
222 LSKYPTTHACREYKFNPLITKTCYREDNVQLEDVSMFLKRSQGTVPVAGVLSPPD 281
169 YLSLQDRYPPIASVWKTLDKDNFSITPDLIHDLGHVPMILHPSSEFPINNRLFTKY 228
282 FLAGLAVRVFCTQYIRHGSDDLTPBPTCHELGHVPLADKXFAQSOEIG----- 335
229 IEKVQALPSKKQRIQTLSNLIAIVRCFWFVESGLIENHGRAYGAVLISSPOELGHA 288

RESULT 1
US-10-154-674-8
;; Sequence 8, Application US/10154674
;; Publication No. US20020192694A1
;; GENERAL INFORMATION:
;; APPLICANT: Yu, Xuanchuan
;; APPLICANT: Miranda, Maricar
;; APPLICANT: Hu, Yi
;; TITLE OF INVENTION: No. US20020192694A1 Human Hydroxylases and Polynucleotides Encc
;; FILE REFERENCE: LEX-0352-USA

```

Db 336 ---LASIGASDEVDQKLT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 385
Qy 289 FIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE-----L 329
Db 386 LSKACVAFDPKTTTCLOECLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 445
Qy 330 TSKLEWMLDQGLLESI 345
Db 446 TQSIIEILKDRSIEINV 461

RESULT 3
US-10-154-674-4
; Sequence 4, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020192694A1el Human Hydroxylases and Polynucleotides Encc
; FILE REFERENCE: LEX-0352-USA
; CURRENT APPLICATION NUMBER: US/10/154,674
; PRIOR FILING DATE: 2002-05-23
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 486
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-4

Query Match 11.5%; Score 216.5; DB 9; Length 486;
Best Local Similarity 26.2%; Pred. No. 5.4e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

Qy 117 LKWSYCPFRFDYLEAFGLSDFLDH-----QAVIKFFFELETHFSYYPVSGFVAPHQ 168
Db 223 LSKLYPTHACREYLKKNFLLTKYCGYREDNVQLEDSVNFELKERSGFTVRPVAGYLSPRD 282
Qy 169 YLSILODRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSSEFFINMGRFLTQV 228
Db 283 FLAGLAYRVPHCTQYIRHGSPLYTPEDTCHELLGHVPLLDAPKPAQFSQEIG----- 336
Qy 229 IEKQVALPSKQRIOTLOSNIILAVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHA 288
Db 337 ---LASIGASDEVDQKLT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 386
Qy 289 FIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE-----L 329
Db 387 LSKACVAFDPKTTTCLOECLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 446
Qy 330 TSKLEWMLDQGLLESI 345
Db 447 TQSIIEILKDRSIEINV 462

RESULT 4
US-10-154-674-2
; Sequence 2, Application US/10154674
; Publication No. US20020192694A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020192694A1el Human Hydroxylases and Polynucleotides Encc
; FILE REFERENCE: LEX-0352-USA
; CURRENT APPLICATION NUMBER: US/10/154,674
; PRIOR FILING DATE: 2002-05-23
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 8

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 490
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-154-674-2

Query Match 11.5%; Score 216.5; DB 9; Length 490;
Best Local Similarity 26.2%; Pred. No. 5.5e-13;
Matches 67; Conservative 42; Mismatches 104; Indels 43; Gaps 4;

Qy 117 LKWSYCPFRFDYLEAFGLSDFLDH-----QAVIKFFFELETHFSYYPVSGFVAPHQ 168
Db 227 LSKLYPTHACREYLKKNFLLTKYCGYREDNVQLEDSVNFELKERSGFTVRPVAGYLSPRD 286
Qy 169 YLSILODRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSSEFFINMGRFLTQV 228
Db 287 FLAGLAYRVPHCTQYIRHGSPLYTPEDTCHELLGHVPLLDAPKPAQFSQEIG----- 340
Qy 229 IEKQVALPSKQRIOTLOSNIILAVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHA 288
Db 341 ---LASIGASDEVDQKLT-----CYFTTIEFGLCKQEGQRAYGAGLLSSIGELKHA 390
Qy 289 FIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE-----L 329
Db 391 LSKACVAFDPKTTTCLOECLITTFQEAIFYSESFEAEKMRDFAKSIITRPPSVYFNPY 450
Qy 330 TSKLEWMLDQGLLESI 345
Db 451 TQSIIEILKDRSIEINV 466

RESULT 5
US-09-205-658-313
; Sequence 313, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: O'G, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 313
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-205-658-313

Query Match 10.9%; Score 206.5; DB 10; Length 532;
Best Local Similarity 25.2%; Pred. No. 5.6e-12;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

Qy 104 RNLM---YLLSSRFSRLKSYCPFRFDYLEAFGLSDFLDH-----QAVIKFFFELETHF 155
Db 251 RKTWGIIVKUR---ELKHGKACKQFLDNFELLERHCYSENNIPQLEDICKFLKAKTGF 307
Qy 156 SYYPVSGFVAPHQYLSILODRYPPPIASVMRTLDKDNFSLTPDLIHDLGHVPHLLHPSFS 215
Db 308 RVRPVAGYLSARDFLAGLAYRVFCTQYVRHADPTTPEPTVHELHGRNALFADDDFA 367
Qy 216 EFTINMGRFLTQVIEKVOALPSKQRIOTLOSNIILAVRCFWFTVESGLI----- 265
Db 368 QFSQEIG-----LASLGASEEDLKLTATL-----YFFSIEFGLSDDAADSPVK 411

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 4

QY 266 ---ENHEGRKAYGAVLISSEPOLIGHAFIDNRVLPLELDIOIRLPNTSTPOETLFSIRH 322
DB 412 ENSNHERKRYGAGLISAGLCHAVGSGTIRRPDPKRVNOEGCITTFQSAFYRN 471
QY 323 FDELVELTSLKEMMDQGLLESIPLYNOEKLSGFEVL 360
DB 472 FEBAQO---KLRMFTNMKRPYRIN--PYTESVEVL 504

RESULT 6

US-09-801-368-354
Sequence 354, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Soile
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jelt
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
PRIOR FILING DATE: 2001-03-07
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 354
LENGTH: 1427
TYPE: PRF
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-354

Query Match 5.0%; Score 95; DB 10; Length 1427;
Best Local Similarity 18.8%; Pred. No. 1.1; Indels 204; Gaps 25;
Matches 98; Conservative 65; Mismatches 155;

QY 12 YIIKIALKTRQSLSEFPONSGLOFAYSTPYRYRII-----LQENKEKQALAHKCI 65
DB 214 YILE---KLIPTMTNHYNSQOL-RTWKRQISYFLKILGNCYSILRLINE---IFHMLV 266
QY 66 SILEFFKNLFF---VHLTSLSKNORECSTDMAVST-----PFF----- 102
DB 267 EFINKEWNEFEFLPLSLHILMIFWMDICQIDINAVPAATITTSQKEPFLVTKITDMLHK 326
QY 103 -----NNLWYRLSSRFSL-----WKS 120
DB 327 YIIVSSSKMINDENYIINDIKONKIKLITLISLILKIFQESQLEVFIFPITSNMEI 386
QY 121 YCPRRF-----LDYLEAFGLSDPLDHOAVIKFPELEHHSYYPVSGFVAP 166
DB 387 YKELLEIYVSNADPTNONSIMKKKLLISYRNESLKNSSSR-----NYIMSNSN 437
QY 167 HOYLSLQDRYFPFIASVM-----RTLDKN-----FSLTPDLIHLIGHVP 207
DB 438 DFOITVITCKQFKLSCIDQNCITDQFTKLID-DNPEFPMFTYVDONPLTMKIIQLIL 496
QY 208 WLLAPS--FSEFFINNGRLFTVIEKVOALPSKQRIQTLQSNL--IAIVRCFWFTVSG 263
DB 497 WSIHPSRQDFHYESN-----OLVAKLLL-----RINSTDEDLHEFOIEDAIVSLVFO- 544
QY 264 LIENHEGRK-----AYGAV-----LISSP-----QELGHAIFID-- 291

DB 545 LAKFNSQKRVSVWMPSLYRLNLTITTYGIKVPYIRKLISSGLLYLQDSNDKRVH 604
QY 232 ---NVRVLPLELDIOIRLPNTSTPOETLFSIRHDELVELTSLK----- 333
DB 605 LINDKISPLMKQYMMVLRNWEVDYKYEITFNPQVLEITTEQIKRILISNDITLQLS 664
QY 334 -----EMMLD---QGLLESIPLYNOEKYISGFEVL 361
DB 665 KTLPLSIKIMVAEWYLSHLCSGLISV---NRTVLKIKFKIPC 703

RESULT 7

US-09-817-774-23
Sequence 23, Application US/09817774
Patent No. US2002012011A1
GENERAL INFORMATION:
APPLICANT: CHOE, Sunghwa
APPLICANT: FELDMANN A., Kenneth
TITLE OF INVENTION: Dwf5 MUTANTS
FILE REFERENCE: 2225-0020 / 91020.002
CURRENT APPLICATION NUMBER: US/09/817,774
PRIOR FILING DATE: 2001-03-26
PRIOR APPLICATION NUMBER: 60/192,202
NUMBER OF SEQ ID NOS: 45
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 23
LENGTH: 427
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: LBR-RAT
US-09-817-774-23

Query Match 4.9%; Score 92; DB 10; Length 427;
Best Local Similarity 20.4%; Pred. No. 0.41;
Matches 87; Conservative 53; Mismatches 138; Indels 148; Gaps 19;

QY 41 PYSYRIILOENKEKQAL-----ARRKCSILEFFKNLFPVHLTSLSKNQ 86
DB 35 PACVFPULLQCAQKDBGLQFPPLPALRELMENAVCGVILWFGLALF-SLLPYGK-- 91
QY 87 REGCSTDMAYSTPFRN-RULWYRLSSRFSIMKSCYCRFPYLDYLEAFGLSDPLDHOAV 145
DB 92 -----VYEGIDPLVDGRKLYRL-----NGLYAPILISAAY 121
QY 146 -----IKFPELEHHSYYPVSGFYAPHQYISLQDRYFPFIASVMRTLDKNSFLTPDI 198
DB 122 GTAVFMDIELYIYTHFLQPALAIV-----FSVLSYLYIYARSLKVPRDELSPASSGNA 176
QY 199 IHDLLGHVPMILHPSFS---EFFINM--GRFTVIEKVOALPSKQRIQTLQSNL----- 247
DB 177 YDFF--IGRELNRIGAFDLKFCFLRGILGWVIVLWMLLEMKVQERSAPSLAMTL 234
QY 248 ---NLIAIVRCFWFTVSGGLENN---HEG---RKAYGAV-----LISSP 283
DB 235 VNSFQLLYVDALWF--EALTLTMDIHDGFWLAFGDVWVPFTYSLOAFYLVNHQ 292
QY 284 ELGHAIFIDNRVLP-----ELDQIRLPNTSTPOETLFS- 319
DB 293 DLSWPLTSVIALKLCQYVIFRANCASQKNAFKNPTDKLAHLKTIIP--ISTKSLIVSG 350
QY 320 ---IRHDELVELTSLKEMMDQGLLESIPLYN-----QEKYLSG 356
DB 351 WKGFRHNPYIGDLIMALAWSLPCGNHILPFIYIYFTALLIHRARDHOCRRKIGLA 410
QY 357 FEVLQ 362
DB 411 WEKYQ 416

RESULT 8


```

US-09-870-759-128
; Sequence 128, Application US/09870759
; Patent No. US20020177551A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 128
; LENGTH: 4563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-870-759-128

Query Match 4.7%; Score 92; DB 9; Length 4563;
Best Local Similarity 20.7%; Pred. No. 10; Indels 68; Gaps 12;
Matches 61; Conservative 48; Mismatches 118;

QY 3 YCERLDPKYLKALKRQSLFQNSQSLORAYSTPYSYR---IIIOKENKEKQAL 59
DB 4211 YREELCTMFIREFVGTWLSQVSKVNGSEIL-----FSIQDLVITLFFE-----L 4257
QY 60 ARHKCISILEFFKNLLFVHLLSLGKNGRGCGSTDMVAVSTPFFNRLW-----YLLSS 113
DB 4258 RHKLIDVISMYREL-----LKDLSEKAEVFKAIOSLKTTEVL-RNIQDLLQIFQIED 4312
QY 114 RFSLWKSYPFRFFDYL--BAFGLSDFLDHOAVIKFPELETHFSYYPVSGFVAP----- 166
DB 4313 NIKQLKEMKFTYLYNQIDEINTIFDIYI--VFELKLENCLNLKHKFNEFIQNELQEA 4370
QY 167 -----HQYLSLQDRYFPFIASVMRT-----LDKDNFSLTPDLI----- 199
DB 4371 SQELQIQHQYIMALREYFDPISVGTWTKYVELEEKIVSLIKNLVALKDPHSEYIVSAS 4430
QY 200 ---HDLGLHPWLPHSPSFPFNMGLFTKVIKVOALPSPKKQRIQTLQSNLIA 251
DB 4431 NFTSLSQSSQVEQLHNIQEVLSILTDPCGKGKIEAELSAQAQEI--IKSQAIA 4483

RESULT 9
US-09-746-491-37
; Sequence 37, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-621
; CURRENT APPLICATION NUMBER: US/09/746,491
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: USN 60/171,329
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 37
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-746-491-37

Query Match 4.7%; Score 88.5; DB 10; Length 511;
Best Local Similarity 21.9%; Pred. No. 1.1; Indels 131; Gaps 19;
Matches 84; Conservative 59; Mismatches 109;

QY 25 SLFFQNSQSLORAYSTPYSYRRIIOKENKEKQALRKHCISILEFFKNLLFVHLLSLK 84
DB 9 SKLINSQQLLYQEYSD-----VVLNKEIQSQORLE-----SL 40
QY 85 NREGCGSTD---MAVVTSTPFFNRLMYRLLSRFSLM-----KS 120
DB 41 SETPGSSPRQPRKALYSSSEYLQRL---SMASGSLMQEIPVVRNNTVLLSMTHEDQKL 97
QY 121 YCERFELDYLEAFGL--LSDFLDHOAVIKFPELETHFSYYPVSGFVAPHQYL-SLLQDRY 177
DB 98 QEVKFEIIVSEASYLSRLNIAVDH-----FQLSTSLR---ATLSNQEHQWLFSLRLQDVR 148
QY 178 FPIASVMRTDK--DNFSLTPDLIHDLLGH-----VPMILHPSFSE---FFINMGR 223
DB 149 DVSAFSLDLENFENNIFSFQVCDVNLNAPDFRVRVLPYVTNQTQERTFQSLMNSNS 208
QY 224 LFTKVIKVOALP-----SKK-----ORIQTLQSNLIAIVRCFWFTVESGLIENHEGR 271
DB 209 NFREVLEKLESDFVQCRSLKSLFILPQRTIRKLKLLQNLK---RTQPGSSEAEAT 264
QY 272 KAYGAVLISSPOELGHAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRHFDELVELTS 331
DB 265 KAH-----HA-----LEQLIR-----DCNNVQSMRTEELIYLSQ 295
QY 332 KLEWMLDQGLLESIPLYNQEKYL 354

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 6

Db 296 KIEPE-----CKIFPLISQSRWL 313

RESULT 11

US-09-746-491-39
Sequence 39, Application US/09746491
Patent No. US20020137202A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: No. US20020137202A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 15966-621
CURRENT APPLICATION NUMBER: US/09/746,491
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: US91 171,329
NUMBER OF SEQ ID NOS: 72
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 39
LENGTH: 550
TYPE: PRT
ORGANISM: Homo sapiens
US-09-746-491-39

Query Match 4.7%; Score 88.5; DB 10; Length 550;
Best Local Similarity 21.9%; Pred. No. 1.3; Indels 131; Gaps 19;
Matches 84; Conservative 59; Mismatches 109; Indels 131; Gaps 19;

Qy 25 SLFFQNSQIQRAISTPYRYRIILQENKQALARKKCSILEFFKALLFVHLISLSK 84
Db 40 SKLINSQQLLYOEYSD-----VINKKEIQSQQRLE-----SL 71
Qy 85 NQREGCSTD---MAVSTPPFNRLMYRLSSRPSLW-----KS 120
Db 72 SETPGSPSPQPRKALVSESYQRL---SWASSGSLMOELPVVNSTVLLSTHEDQKL 128
Qy 121 YCFRFLDYLEAFGL--LSDPLDHOAVIKFELETHFSYVSGFVAHQYL--SLQDRY 177
Db 129 QEVFELIVSEAGYLSRLNIAVDH-----FOLSTSLR--ATLSNQEHQWLPFRSQDR 179
Qy 178 FPLASVWRTDK--DNFSLTPDLIDHLLGH-----VPLLHPSPSE---FPIWGR 223
Db 180 DVATFSLDLEENFENNIFSPQCVVLTNHPDRRVLYPVYNGTYQERTFQSLNKS 239
Qy 224 LPTKVIKQVALP-----SK-----QRTQSQNLIAVRCFMTVESGLIENHEGR 271
Db 240 NFRVLEKLSDPVQRLSKSFLPLPFRITRLKLLQNLK---RTQSSSEAEAT 295
Qy 272 KAYGAVLISPOELGHAFLDNVRLPLEDQIIRLPFTSTPOETLFSIRHPELVELTS 331
Db 296 KAH-----HA-----LEQLIR-----DCNNNVQSMKREELIYSQ 326
Qy 332 KLEWMLDOGLLESIPLYNOEKYL 354
Db 327 KIEPE-----CKIFPLISQSRWL 344

RESULT 12

US-09-981-421-4
Sequence 4, Application US/09981421
Patent No. US20020098185A1
GENERAL INFORMATION:
APPLICANT: Sims, John E.
APPLICANT: Mohler, Kendall M.
TITLE OF INVENTION: METHODS FOR TREATING IL-18 MEDIATED DISORDERS
FILE REFERENCE: 3086-A
CURRENT APPLICATION NUMBER: US/09/981,421
PRIOR FILING DATE: 2001-10-17
PRIOR APPLICATION NUMBER: US 60/241,408
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.1

; SEQ ID NO 4
; LENGTH: 541
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-981-421-4

Query Match 4.6%; Score 87.5; DB 10; Length 541;
Best Local Similarity 21.3%; Pred. No. 1.5; Indels 95; Gaps 19;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

Qy 9 DPXYI---IKIALKQSLSLFFQNSQIQRAISTPYRYRIILQENKQALARKK 64
Db 34 EFPYIKQSCSLAHEITTKSWYKSSQCEHVELNPFSSRIAL-----HQC 81
Qy 65 ISILEFF-----KNLLEFVHLISLSKNQREGCSTDMAVVS---TPFN-- 103
Db 82 --VLEFWEVLENDTGYFFQMKNYTQKMLNVIIRNKHSCPTERQVTSKIVEYKFFQIT 139
Qy 104 -RNLWRLSSRFELMKSQCFRFLDYLEAFGLSDPLDHOAVIKFELETHFSYVSG 162
Db 140 CENSYQTLVNSTSLYKN--CKKLLLENNK-----NPTIKNAEPEDQ--GYSCVH 187
Qy 163 FVAPHQYSLQDRYPPLASVWRTDKDNFSLTPDLIDHLLGHVPMILHPSFSEFLNMG 222
Db 188 FL--HH-----NGKLFVITK-----TFNIT--IVDRSNIVPVLLGPKLHVAVELQ 230
Qy 223 RLFTKVIKQVALPSSKQRTQSQNLIAVRCF--FVESGLIEN--HEGRKAYGAVLS 280
Db 231 K--NRLNCSALNEDVLI-----YMFGEENGSDPNTIHEKE---NRIM 270
Qy 281 SPQELGHAFLDNVRLPLE 299
Db 271 TPEGKMAH---SKVLRIE 285

RESULT 13

US-10-157-447-2
Sequence 2, Application US/10157447
Patent No. US20020143155A1
GENERAL INFORMATION:
APPLICANT: Parnet, Patricia et al.
TITLE OF INVENTION: Receptor Designated 2p1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/157,447
FILING DATE: 28-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/578,178
FILING DATE: <Unknown>
APPLICATION NUMBER: 09/110,618
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Anderson, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2619
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-157-447-2

Query Match 4.6%; Score 87.5; DB 12; Length 541;
Best Local Similarity 21.3%; Pred. No. 1.5;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

QY 9 DPKYI---LKALKRQSLSPFQNSQSLQRAYSTPYYYRIILOKENKEKOALAHKC 64

Db 34 EPYLUKHCSCSLAHEIETTTKSWKSGSQEHVELNPRSSRIAL-----HDC 81

QY 65 ISILPFF-----KLLFVHLLSLKNQREGCTDMAVVS-----TPPFN-- 103

Db 82 --VLEFWPELNDTGSYFFQMKNYTKQKLVIRRNKHSCTERQVTSKIVEVKFFQIT 139

QY 104 -RLNMYRLSSFSLSKWSYCPREFLDYLEAFGLSDFLDHOAVIKFPELETHFSYYPVSG 162

Db 140 CENSYQTLVNSTLYKN-CKLLENK-----NTIKNAEFEDQ-GYYSVCH 187

QY 163 FVAPHOYLSLQDRYPPIASVMRTLDKDNFSLTPDLIHDLLGHVPWLLHSPSEFFINMG 222

Db 188 FL--HH-----NGKLFNITK-----TFNIT--IVDRSNIVFVLGPKLNHVAELG 230

QY 223 RLFTKVIKVOALPSKQRIOTLQSNLIAIVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280

Db 231 K---NVLNLCGALLNEEDVI-----YNNFGEENGSDFNTHSEKE---MRIM 270

QY 281 SQPELGHAFIDNRYVLPLE 299

Db 271 TPEGKWAH-----SKVLRIE 285

RESULT 14

US-09-864-761-43207
Sequence 43207, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 43207
LENGTH: 978
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008174.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: SWISSPROT HIT: Q62209, EVALUATION 5.00e-03
US-09-864-761-43207
Query Match 4.4%; Score 84; DB 10; Length 978;
Best Local Similarity 17.9%; Pred. No. 7.4;
Matches 53; Conservative 42; Mismatches 77; Indels 124; Gaps 10;
QY 8 LDPKYLKALKRQSLSPFQNSQSLQRAYST-----PYSYRIILOKENKEKOALA 60
Db 639 VQDTYKLAIDPKORELSFVNSVLEEVISELLCKILYAFSHMLVTENPDVKLKT 698
QY 61 R-----HKCIS-----ILEFFKNLLFV 77
Db 699 RIVTTLVNSIVLETTSEILVADNFKNLCPSERYKENVKIVNSVYGVKVLDOYKSLIQI 758
QY 78 HLLSLKNQREGCTDMAVVSTPFFNRLNWLSSRFLSKWSYCPREFLDYLEAFGLIS 137
Db 759 HEVIOQ-----DTICFGKRIYLLLEIY-----DY-QVOSLVS 791
QY 138 DFLDHOAVIKFPELETHFSYYPVSGFVAHQVLSLLQDRYPPIASVMRTLDKDNFSLTP- 196
Db 792 G-----ELESSYSYPOADNI-----IRNLNITKDSHALPPY 825
QY 197 -----DLIHDLLGHVPWLLH-----PSFSEFFINMGRFTKVIKV 232
Db 826 ITVLPHELLEDWVRLGHVPSTHTTENELKEKFPDPDEFVEAASKLTDEIIKEI 881
RESULT 15
US-09-741-669-385
Sequence 385, Application US/09741669
Patent No. US20020022718A1
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Alllyn
APPLICANT: Chisen, Kari L.
APPLICANT: Chisen, Judith W.
TITLE OF INVENTION: Genes identified as required for
TITLE OF INVENTION: proliferation of E. coli
FILE REFERENCE: ELITRA.009A
CURRENT APPLICATION NUMBER: US/09/741,669
CURRENT FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: US 60/173005
PRIOR FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 385
LENGTH: 450
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-669-385

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 8

```
Query Match 4.44; Score 83.5; DB 10; Length 450;
Best Local Similarity 21.74; Pred. No. 2.9; Mismatches 57; Gaps 12;
Matches 47; Conservative 37; Indels 102; Gaps 67; Gaps 12;

QY 106 LWRLLSRFSLKSYCRFLDYLKAGLLSDFLDHOAVIKFELEHTFSTYPSGVA 165
DB 215 LFGIIRGNISL-----MLAAGV-----GVGMFWLAKRYGMWNSGDEA 257
QY 166 PFOYISLADRYFP--IASVWRTLDKONFSLTPDLIHLLGHVWMLHPSFSEFFIMG 222
DB 258 FTYFLVLTTRDTPSPWENALLLOVDNIDFGGLAPVRDPEYVFPISWMLPGPSMVA 317
QY 223 RLFT-KVLEKVALPSKQRIQTLQSNLIA-----YVRCFWFVESGLI 265
DB 318 NYFTWEVLNNHSGLAIS-----PTLIGSLVWVGALFPLGAIYVGLIKFMDLYELG-- 371
QY 266 ENHEGRKAYGAVALISSPQELGHAF---IDNRYVLPLE-LDQIIRLPFTSTPQRTLP 320
DB 372 -NREPRRYKAAL-----HSFCFGAIFPMVTLAREGIDSV-----SRVVFPI 413
QY 321 RHDEBELVELSKLEWMLDO-GLL 342
DB 414 VVFGACIMIAKLIVWLPESAGLI 436

RESULT 16
US-09-801-574-22
; Sequence 22, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang
; APPLICANT: Wang, Peijiang
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399,2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 835
; LENGTH: 935
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-801-574-22

Query Match 4.44; Score 83.5; DB 10; Length 835;
Best Local Similarity 24.44; Pred. No. 6.7;
Matches 47; Conservative 26; Mismatches 47; Indels 73; Gaps 10;

QY 210 LHPFSEFFINMRLFTYVIRKVALP--SKQRIQTLQSNLIAVRCFWFVESGL-- 264
DB 155 LHPSSSTFLHNVGLLENQFIRKRFPSDLAKNEK---OSNLKDSIRDF---EANLVVC 206
QY 265 IENHEGR-KAYGAVALISSPQELGHAFIDNRYVLPLELDQIIRLPFTSTPQRTLP 323
DB 207 ISNEKGERNVAREVIDISKP---GFCF-----PFTNYPEDSGVDVADL 246
QY 324 DELVE-----LTSKLEW-----MLDQGL-----LES 344
DB 247 NDITKLFSPVLETHCHENGLENHEHMKYTYLLPELWGLNPGNTCYINVLVGLS 306
QY 345 IPIYNOKYLSGF 357
DB 307 IPIFINDLPNGCF 319

RESULT 17
US-09-864-49017
; Sequence 49017, Application US/09864761
```

```
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenming
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmtca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 49017
LENGTH: 995
TYPE: PRT
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AC007282.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6
OTHER INFORMATION: EST HUMAN HIT: A1023682.1, EVALU8 8.00e-72
OTHER INFORMATION: SWISSPROT HIT: P30622, EVALU8 1.00e-06
US-09-864-761-49017

Query Match 4.44; Score 82.5; DB 10; Length 995;
Best Local Similarity 20.94; Pred. No. 11;
Matches 70; Conservative 45; Mismatches 117; Indels 103; Gaps 15;

QY 78 HILSLSNRREGCTDM-----AVSTRPFRN-----NL-----MYRL 111
DB 110 HLISEVNSKSSSGTVDHYIMKQIFAPISLEIEYKESPSTPMNLQPLTPKXSL 169
QY 112 SGRPSLMKSYCPFRFL-----DYLEAFGLSDPLDHOAVIKFPEL--ETHFSTYPS 161
DB 170 SSHLPHEENADEIELPQRSATSOIIQAFPI--DTLLSGIILKVELDKENHKSLLGT 227
QY 162 GFVAHQYISLADRYFPYASVWRTLDKONFSLTPDLIHLLGHVWMLHPSFSEFFIM 221
```

Db 228 GITSKGNLQDSQYSEIRSETEPLSEQSIPIPKDTTSV-----SRAEFIQED 277
QY 222 GRLFTK-----VIEKQALPSKKORIOTLQSNLIAVRCFWFTVESGLIENHEGRKAYG 275
Db 278 QNMFQDSSVYSYANKELYPRNGQRLCKDKNDLSLSTNLMDKLSSEDE----- 331
QY 276 AVLISSPQELGHAFIDNV-----RVLP-LELDQIIRLPNTSTPQETLFSIRH 322
Db 332 -IMLKS-----FLKNIFNVFFKYNHSSRRQPEKELERLIQPSFTSDT-----EH 375
QY 323 FDELVE-----LTSKLEWMLDQGLLES 344
Db 376 LEELOEDFKADKLDKRPILSPKLRVFLSE-LSES 409

RESULT 18
US-09-864-761-43061
; Sequence 43061, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 43061
LENGTH: 1202
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO AC007282.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.81
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.72
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 0.73
OTHER INFORMATION: EST HUMAN HIT: A1023682.1, EVALUATE 1.00e-71
OTHER INFORMATION: EST HUMAN HIT: A1023682.1, EVALUATE 1.00e-71
OTHER INFORMATION: SWISSPROT HIT: P30622, EVALUATE 2.00e-06
US-09-864-761-43061
Query Match 4.4%; Score 82.5; DB 10; Length 1202;
Best Local Similarity 20.9%; Pred. No. 14;
Matches 70; Conservative 45; Mismatches 117; Indels 103; Gaps 15;
QY 78 HLLSLNQREGCSDM-----AVVSTPFNR-----NL-----WYRL 111
Db 317 HLLSEVNSKSGSGTVDYIMRQITAFIPSELEIEVSEKSPETPMNLENQLPTWKRS 376
QY 112 SSRFSLMKSYCPFFL-----DYLEAFGLSDFLDHOAVIKFEL--ETHFSYYPVS 161
Db 377 SSHLFHEENADEIELPQPRSATSOIIQAFPI--DTLESGITKVILOKHHKSL 434
QY 162 GFVAPHOYLSLQDRYFPIASVNRVTLDKDNFSLTFDLIHLHLGHVWMLHPSFSEFFINM 221
Db 435 GITSKGNLQDSQYSEIRSETEPLSEQSIPIPKDTTSV-----SRAEFIQED 484
QY 222 GRLFTK-----VIEKQALPSKKORIOTLQSNLIAVRCFWFTVESGLIENHEGRKAYG 275
Db 485 QNMFQDSSVYSYANKELYPRNGQRLCKDKNDLSLSTNLMDKLSSEDE----- 538
QY 276 AVLISSPQELGHAFIDNV-----RVLP-LELDQIIRLPNTSTPQETLFSIRH 322
Db 539 -IMLKS-----FLKNIFNVFFKYNHSSRRQPEKELERLIQPSFTSDT-----EH 582
QY 323 FDELVE-----LTSKLEWMLDQGLLES 344
Db 583 LEELOEDFKADKLDKRPILSPKLRVFLSE-LSES 616

RESULT 19
US-09-815-242-12990
; Sequence 12990, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 10

SEQ ID NO 12990
LENGTH: 909
TYPE: RT
ORGANISM: Staphylococcus aureus
US-09-815-242-12990

Query Match 4.3%, Score 82; DB 10; Length 909;
Best Local Similarity 22.4%; Pred. No. 10;
Matches 87; Conservative 66; Mismatches 144; Indels 92; Gaps 22;

QY 11 KYLLKIALKLRQSLSPFQNSQSLQRAYSTP---YSYRI--ILQKE-----NKEQA 58
DB 512 KFOYSITKQKQYRTDLLFQFQNDISIKQYTVENLLINAGYQINLQOSQTTIYVINOQKVI 571
QY 59 LARHKCISILFEFFKNLLFVHLISLS---KNQRE-GCSTDMAVSTPFFNRNLMYRLSSR 114
DB 572 ----KTIPQNHIDNTTQOHRQALSWVKNERQAGATTD-----TFPGINKW--LIPIG 619
QY 115 FSLMKSCYCRFLDYLEAFGLSDPLDHOA--VIKFE--LETHSYYPVGFAVAPHOY 169
DB 620 TSPFK-----GILA--IDYQSQVINPYDASILESMLESLA-----VEN 658
QY 170 LSLIDRYFPFASVWRTDKONF--SLTPDL--IHLLGHVPMILHPSFEFFINMRL 224
DB 659 VILLKQTRFESMLQAEKOLTHSNFLRSISHDRTPLTTIMGNLILVSHKD-----709
QY 225 FTKVIERQVALP---SKQRIQTLQSNLIAVRCFWTVE-----SGLIEN---HE 269
DB 710 -MSIIEKQLLVHSFQESQYLYLVNTNLSTKLQSSNVQIKQPYLSELVEIDMILE 768
QY 270 GRKAYGAVLISSPOELGAFIDNVAVLPELIDQIIRLPFNTSTPOETLESIRHPD-ELVE 328
DB 769 RRLKKRITVSSSVNLQFHHDSKILQALFNLIENAVKATSTDTKINLSIRVASYEOIE 828
QY 329 LTSKLEMLDQGLLESIPLYNOEKYLSGF 357
DB 829 FA-----VIDEG--FGISLEQOKIFEPF 850

RESULT 20
US-09-833-790-366
Sequence 366 Application US/09833790
Patent No. US20020068288A1
GENERAL INFORMATION:
APPLICANT: Lodee, Michael J.
APPLICANT: Wang, Tonglong
APPLICANT: Secretist, Heather
APPLICANT: Mohamach, Radooh
APPLICANT: Indrias, Carol Y.
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.512
CURRENT APPLICATION NUMBER: US/09/833.790
CURRENT FILING DATE: 2001-04-11
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 366
LENGTH: 818
TYPE: PRT
ORGANISM: Homo sapiens
US-09-833-790-366

Query Match 4.3%, Score 81.5; DB 10; Length 818;
Best Local Similarity 19.8%; Pred. No. 10;
Matches 57; Conservative 39; Mismatches 93; Indels 99; Gaps 12;
QY 95 AAVSTP-----FNRNLMYRLSSRFSJWS-YCPREFLDYLEAFGLSDF 139
DB 27 AFVTTMCCPSRSSMLTGKRYVNNHVVYNNENCSPPSQAMHEPRTFAVYLLNTGYRTAF 86
QY 140 LDHQAIVIKFELETHFSYYPVGFAVAPHOYLSLQDRYFPIASVMT--LQDNQSLTPD 197

DB 87 FG-----KYLN-BYNSYIP-PGW---REWGLIKNSRFYNTVTCNGIKKHGPDYAKD 136
QY 198 LIHDL-----GHVPMILHPSFEFFINMCR-----223
DB 137 YFTDLITNESINFKSKMYPRHRYVNVVISHAAPHPEDSAIQFSKLYPNMSQHTTSEY 196
QY 224 -----LFTKXIEKQVQLPESKQRIQTLQSNLIAVRCFWTV 260
DB 197 NYAPNDKHWIWOYTPMLPIHMEFTNLQ-----RRQLTMSVDSVRLNMLV 248
QY 261 ESSLIEHNEGRKAYGAVLISSPOELGAFIDNVAVLPELIDQIIRLPF 308
DB 249 ETGELEN-----TYIITYTDHGHYHIGQGLVKGSMEDPDD--IRVPF 289

RESULT 21
US-09-815-242-13312
Sequence 13312 Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Olsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Grant, J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13312
LENGTH: 1216
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13312

Query Match 4.3%, Score 81.5; DB 10; Length 1216;
Best Local Similarity 21.6%; Pred. No. 17;
Matches 69; Conservative 46; Mismatches 123; Indels 81; Gaps 15;
QY 7 TLDPKYLKIALKLRQSLSPFQ-----NSQSLQRAYSTPYSYRIILO--KENKERO 57
DB 556 TGEKRLVIRKEILKLBQEKVAKFEIALLTSRSRNDQILLALBEYIGIPVKTDOEONNYAQ 615
QY 58 ALARHKCISILFEFFKNLL-----FVYL-----LSLK-----84
DB 616 SLEQVQWLDLRYIHPLODYALVALMKSPPFGDEDELARLSLQKAEKDHENLYEKLV 675
QY 85 NQREGSTDMAVST-----PFRNLMYRLSSRFSJ--WKSYPREFLDYLEAF 133
DB 676 NAOQWASQSGLIHTLAELAKLQFMDILASWRLVAKTHSLYDLIWKIYNDRFYDYV---732
QY 134 GLISDFLDHQAIVIKFELETHFSYYPVGFAVAPHOYLSLQ--DRYFPIASVMTLDD 190

```

Db 733 GALPNGPARQA--NLVALALRADQFEKNFGLSRFIRMIDQVLEAQHDLASVAVAPPKD 790
Qy 191 NFSJTPDLIDHLG-HVPMLHPSPSEFFINNGRLFTKVIKQVQALPSKKORIQTLOSNL 249
Db 791 AVELM--TIHKSGLFFPV-----FILNMQDFNKQDSMSEVILSRQ---NGLGVKY 838
Qy 250 IAIVRCFWFTVESGLIENH 268
Db 839 IA-----KMETGAVEDH 850

RESULT 23
US-08-984-090-2
; Sequence 2, Application US/08984090
; Patent No. US20010021502A1
; GENERAL INFORMATION:
; APPLICANT: Swift, Michael R.
; APPLICANT: Athma, Prasanna
; APPLICANT: Li, Alrong
; TITLE OF INVENTION: Predisposition to Breast Cancer by
; TITLE OF INVENTION: Mutations at the Ataxia-Telangiectasia Genetic Locus
; NUMBER OF SEQUENCE: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701-E
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,090
; FILING DATE: 03-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,435
; FILING DATE: 03-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ihnen, Jeffrey L
; REGISTRATION NUMBER: 28,957
; REFERENCE/DOCKET NUMBER: 2323-103
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-783-6040
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-984-090-2

Query Match 4.3%; Score 81.5; DB 8; Length 3056;
Best Local Similarity 18.5%; Pred. No. 60;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17

Qy 10 PKYILKIAL-----KLRQSLSLFFQNSQSLQRAYSTPYVYRIL---QKENKEQ 57
Db 1382 PSHVITAKFAYISNCHTKLSILELSK-----SPDSYQKILLAIACEQAETNN 1431
Qy 58 ALARHKCISILEFFKNLLFVHLLSLKQNBEGCSTDMVSTPFENRLNWLRLSSRFSL 117
Db 1432 VYKGRHLKILYHLFVLSLL-----KDIKSLGGAWA-----FVLRDVIYTLIH---- 1474
Qy 118 WKSVCYRPFPLDY-LEAFGLLSDFLDH--QAVIKFPE--LETHFSYYPVSGFVAP----- 166
Db 1475 YINQRPSCTMDVSLRFSFSLCCDLSQVCPTAVTYCKALENHL--HVIVGTLLPLVYEQV 1532
Qy 167 ---HOYLSLLQDRYPFIASVMRTLDKONFSLTDLIHLLGHV----- 206
Db 1533 EVQKQVLDLL--KYL-----VIDNKDNENLYITIKLDPPDHVVPKDIRITQOKIKYRG 1586
Qy 207 PWLHPSPSEFFINNGRLFTKVIKQVQALPSKKORIQTLOSNLIAIVRCFWFTVESGLIE 266

Db 733 GALPNGPARQA--NLVALALRADQFEKNFGLSRFIRMIDQVLEAQHDLASVAVAPPKD 790
Qy 191 NFSJTPDLIDHLG-HVPMLHPSPSEFFINNGRLFTKVIKQVQALPSKKORIQTLOSNL 249
Db 791 AVELM--TIHKSGLFFPV-----FILNMQDFNKQDSMSEVILSRQ---NGLGVKY 838
Qy 250 IAIVRCFWFTVESGLIENH 268
Db 839 IA-----KMETGAVEDH 850

RESULT 22
US-09-815-242-13698
; Sequence 13698, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13698
; LENGTH: 1216
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-815-242-13698

Query Match 4.3%; Score 81.5; DB 10; Length 1216;
Best Local Similarity 21.6%; Pred. No. 17;
Matches 69; Conservative 46; Mismatches 123; Indels 81; Gaps 15;

Qy 7 TLDPKYLKIALKRLQSLSLFFQ-----NSQSLRAYSTPYVYRILQ---KENKEQ 57
Db 556 TGMRLVKEILLKHOEGVAFKETAITLTSRRNDQILIALSEYIGPIKTYDGEQNNYLQ 615
Qy 58 ALARHKCISILEFFKNLL-----FVHL-----LSLSK----- 84
Db 616 SLEVQVMDLTRVHNPLQDVALVALMKSPMGFDEDLARLSLQAEKDVHENLYEKL 675
Qy 85 NQREGCSTDMVAVST-----PFENRLNWLRLSSRFSL-----WKSVCYRPFPLDYLEAF 133
Db 676 NAQRWASSQGLIHTLAELAKQFQDILASRWLVAKTHSLYDLIMKYNDREYVDV--- 732
Qy 134 GLISDFLDHQAIVKFELETHFSYYPVSGFVAPHOYLSLQ---DRYPFIASVMRTLDK 190
Db 733 GALPNGPARQA--NLVALALRADQFEKNFGLSRFIRMIDQVLEAQHDLASVAVAPPKD 790
Qy 191 NFSJTPDLIDHLG-HVPMLHPSPSEFFINNGRLFTKVIKQVQALPSKKORIQTLOSNL 249

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 12

Db 1587 PSLLEBETH-FLSVYVDALPLTRLEGLKOLRQLHNDQVMDIMRASQONPDQIMV 1645
QY 267 -----NHEGRK-----AYGAVL 278
Db 1646 KLVVNLQJLSKRAINHTEKEVLEAVSSCL 1675

RESULT 24
US-09-801-368-352
Sequence 352, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: Patent in version 3.0
SEQ ID NO 352
LENGTH: 323
TYPE: PRF
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-352

Query Match 4.3%; Score 81; DB 10; Length 323;
Best Local Similarity 23.0%; Pred. No. 3.2; Mismatches 86; Gaps 12;
Matches 56; Conservative 31; Indels 71; Indels 86; Gaps 12;
QY 122 CRRFLDYL-EAFGLSDPL-DHOAV--IKFELTHRSYFVSGFVAPRQYLSLQDR 176
Db 134 CPOYIRTLVSEARTLMPETIPDPPTKTEFEFFYLEESY-----LIVHHPQSLKO-- 186
QY 177 YFFIASVNR-----TLDKNFSITPDLIHDLGHVWMLHPSFSEFFINMGRLLTKVIE 230
Db 187 ---IVQVKOPPOITLSSDDQNCWSLINDSYINDVHLLYPPH----- 227
QY 231 KYQALPSKKORIQTLSNLIIVRCFMTVESGLIENHGRKAYGAVLSSPQELGHAFI 290
Db 228 -----ILVACLFTIT-----SIHKKPTKSSLSAASE----- 256
QY 291 DNVRLVPLELDQIIRLPNTSTPOTLFSIRHPELVELSKLEMMLDGLLESIPLYNQ 350
Db 257 -----AIRDPKSSSSPVQIAIRFMASIVD-----LEVMWDI-IDEQTLIVDH 299
QY 351 -EKY 353
Db 300 WDKY 303

RESULT 25
US-09-765-272-60
Sequence 60, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 474 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-09-765-272-60

Query Match 4.3%; Score 81; DB 10; Length 474;
Best Local Similarity 20.2%; Pred. No. 5.4; Mismatches 113; Indels 76; Gaps 10;
Matches 59; Conservative 44; Indels 113; Indels 76; Gaps 10;
QY 135 LLSPLDHO--AVIKF--FELTHRSYFVSGFVAPRQYLSL-----LDRVPPIASVNR 185
Db 53 LVDKSSKILKLVKIVGFSQNSYKTLPSSTYIAPSTSNPFTIKDEKLFELIQOE 112
QY 186 TLDKNFSITPDLIHDLGHV-----PMLHPSFS--EFTINNG 222
Db 113 LTKAKNLSLQSGIGRUTANELRIIVSEKLSAFNFFNOETKCLTETISFSPVPFANQVG 172
QY 223 RLFTKVIKQVAL--PSKKORIQTLSNLIIVRCFMTVESGLIENHGRKAYGAVL 279
Db 173 ERFANLSDLIDTYKKARSDRVKQASBLR-----RVENELQKNRRKUKOKOKELL 225
QY 280 SSP-----QELGHAFIDNVRLVPLELDQIIRLPNTSTP----- 313
Db 226 ATDNAEERFROKGBLITTFHQVNPNDQVYIINDYTNQPIIMIALDKALTEPQNAQRFFKR 285
QY 314 -QELTFIRHPELVELT-----SKLEMMLDGLLESIPLYNOEKYLSGF 357
Db 286 YQKKEAVKLTDLIEETKATILYVESVETVINDQAGLEIAERRELDIQGF 337

RESULT 26
US-09-895-913A-244
Sequence 244, Application US/09895913A
Patent No. US20020160456A1
GENERAL INFORMATION:
APPLICANT: Kleanchous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean Francois
APPLICANT: Oomen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Genomic
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-244

Query Match 4.3%; Score 81; DB 9; Length 823;
Best Local Similarity 20.7%; Pred. No. 11;
Matches 59; Conservative 44; Mismatches 116; Indels 66; Gaps 14;

QY 17 ALKRWQSLSPFQNSQSLQRAYSTPYVYRIILOKENKEKOALAHKCI-SILEF---F 71
DB 559 SRLRHLAGIPLKSLKLAN-FIAMVMPKNLLNTKYEATRTKLEKGVGAILDFGELGF 617
QY 72 KLLFVHLLSKNORE-----GCSTMAVSTP--FENRLWYLLSSRFLSKSCYCPRF 125
DB 618 KGVLVETTAIVTKSKEVLARSPLNLSIKQKPSYIFDQLPY-----WVIYNAP 668
QY 126 P--LDYLEAPGLSDFLDHQ-----AVIKFPELETHFSYYPVSGFVAPHQYLSL 172
DB 669 FDKVHMQPGLFVDFDRDITNSVLKNGIRVKSRIIDEN-----GKIENIDSY 722
QY 173 LDQRYPIASVWRTLDKDNFSLPDIHLLGHVPMWLLHPSFSEFFINMGLRTKVIK 232
DB 723 IQKEVLSFPKIASFLDRDDVLTNPNTYK-----PRILKKG-KGYVWNGS--VAILIPKN 774
QY 233 QALPSKKQ-----RIOTLOSNIIVRCFWFTV 260
DB 775 PISLSKKQCDYISSVEFRDFYKIARNYQTRTLN-IDSMSCFWFI 818

RESULT 27
US-10-098-979-2
; Sequence 2, Application US/10098979
; Patent No. US2002017207A1
; GENERAL INFORMATION:
; APPLICANT: Wyriad Genetics, Incorporated
; APPLICANT: Sugiyama, Daniel
; APPLICANT: Cimbora, Daniel
; TITLE OF INVENTION: ISG101-INTERACTING PROTEINS AND USE THEREOF
; FILE REFERENCE: 1907.06
; CURRENT APPLICATION NUMBER: US/10/098,979
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/276,259
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US 60/304,101
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: to be assigned
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: to be assigned
; PRIOR FILING DATE: 2002-01-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1993
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-098-979-2

Query Match 4.3%; Score 81; DB 9; Length 1993;
Best Local Similarity 18.0%; Pred. No. 38;
Matches 70; Conservative 61; Mismatches 121; Indels 136; Gaps 19;

QY 30 NSOSLQRAYSTPYVYRIILOKENKEKOALAHK--CISILEFFFKNLLFVHLLSKNOR 87
DB 1114 NSEIQRK-----LNEQOELQDIEKHSTGVASVNLCEVLL-----HDC 1153
QY 88 EGGSTMAVSTPFFNRLN-----WYRL-----LSRFLSKSCYCPRF--FLDYLEA 132
DB 1154 DACATDAECDSIQATRNLDRRWRNICAMSWERELKIEETWRLWOKFLDDYSRFDWLKS 1213
QY 133 FGLLSFLDHQAVI-----KP--PELETHFSYYPVSGFVAPHQYLSLQDRYFFIA 181
DB 1214 SRTAAFPSSSGVIYTVAKEELKKEAFORQVHCL-----TOLELINKQYRELA 1263
QY 182 SVMRTLDKNFSLTPDLHLLGHVPMWLLHPSFSEFFINMGLRTKVIKVOALPSKKOR 241
DB 1264 RENET---DSACSLKQWHE--GNORW-----DNLOKRVTSILRLKHFIFGOREE 1308
QY 242 IQTLOSNIIVRCFWFT-----VESGLIENH---EGRK 272
DB 1309 FETARDSILV-----WLTMDQLTNIEHFSECDVQAKIKOLKAPQOEISLHNKIEQII 1363
QY 273 AYGAVLISSPQELGHAFIDNVRVLELDQIIRLPNTSTPQETLFSI-RHFDDELVELTS 331
DB 1364 AQGEQLIEKSEPLDAAIIEE-----ELDELRY-----COEVFGVRVRYHKKLIRL-- 1409
QY 332 KLEWMLDQGLLESIPLYNOEKYLSGFV 359
DB 1410 -----PLPDDHDLSDREL 1423

RESULT 28
US-09-764-853-836
; Sequence 836, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: FUZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 836
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-853-836

Query Match 4.3%; Score 80.5; DB 10; Length 197;
Best Local Similarity 23.8%; Pred. No. 1.8;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;
QY 197 DLIHLLGHVPMWLLHP-----SFSEFFINMGLRLE---TKVI-EKVQALPSKKQ 240
DB 10 DALYDMIGFPDFILEPKELDDVDVGVGEISEDSFFQMLNLYNFSKVMADQLRKPSRDQ 69
QY 241 RIOTLOSNIIVRCFWFTVESGLI-----ENHEGRKAYGAVLISSPQELGHA 288
DB 70 WSMTPQT-----VNAYVLPKNSIVFPAGILOAPFYARNHPKALNFGGIGVWGHETHA 124
QY 289 FIDNVR 294
DB 125 FDDQGR 130

RESULT 29
US-09-978-295A-526
; Sequence 526, Application US/09978295A
; Patent No. US20020156006A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 14

APPLICANT: Desnoyers, Luc
APPLICANT: Baron, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Shekman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C11
CURRENT FILING DATE: US/09/978,295A
PRIOR APPLICATION NUMBER: 2001-10-15
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-06
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558

PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;
Qy 197 DLHDLGLHVPMLHP-----SFSEFFINWGLF---TKVI-EKVQALPSKKQ 240
Db 464 DAYDMIGFPDFILPEKELDDVDYDGYISEDSFFQNLNLYNFSKAWADQKPPSRDQ 523
Qy 241 RIOTLQSNLIAIVRCFWFTVESGLI-----ENHEGRKAYGAVLISSPQELGHA 288
Db 524 WSMTPQT-----VNAYILPTKNEIVFPAGILOAPFYARNHPKALNFGGIVGVWGHETHA 578
Qy 289 FIDNVR 294
Db 579 FDOGR 584

RESULT 30
US-09-978-697-526
Sequence 526 Application US/099786697
Patent No. US20020169284A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi
APPLICANT: Baker Kevin P.

APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gershten, Mary E.
APPLICANT: Goddard, Audrey J.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC27
CURRENT APPLICATION NUMBER: US/09/978,697
CURRENT FILING DATE: 2001-10-16
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 16

PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/08336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083544
PRIOR FILING DATE: 1998-04-29

PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/08366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084411
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

QY 197 DLHDLGHVWMLHP-----SFSFPIINNGRLF---TKVI-EKVOALFSKKQ 240
DB 464 DAYDMIGFDFILKEKEDVDYDYEISEDSFFQNMNLNYSASAKVMDOLRKPSSRQ 523
QY 241 RIOTLQSNLAIYRCFWFVSESGLI-----ENHGRKAVGAVLISSPOTLGA 288
DB 524 WSMTPQT-----VNAYLPTKNEIVFPAGILOAPFYARNHPKALNFGIGVWGHETLA 578
QY 289 FIDNVR 294
DB 579 FDDQGR 584

RESULT 31
US-09-978-192a-526
Sequence 526, Application US/09978192A
Patent No. US20020177553A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David B.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630P1C9
CURRENT APPLICATION NUMBER: US/09/978,192A
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 18

```

PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/083666
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

QY 197 DLHDLGHVPLHP-----SFEFFINNGRLF--TVI-EKVOALPSKKQ 240
DB 464 DAYDWIGFPDLEKELDVYDGEIYSEDSFQNNLNTYNSAKVAMDGARKRPSRDQ 523
QY 241 RIOTLQSNLIATRCWFVSEGL-----ENHEGRKAYGAVLISPOELGHA 288
DB 524 WSWTFPT-----VNAVYLPYKKEIVFPAGILQAFYVARNHPKALNFGIGVMGHELTHA 578
QY 289 FIDNVR 294
DB 579 FDDGR 584

RESULT 32
US-09-999-832A-526
; Sequence 526, Application US/09999832A
; Publication No. US20020192706A1
; GENERAL INFORMATION:

APPLICANT: Asphenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James.
APPLICANT: Pami, Nicholas P.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC63
CURRENT FILING DATE: 2001-10-24
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
```


Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 20

GENERAL INFORMATION:

APPLICANT: Aekkenazi, Avi
APPLICANT: Baker Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan
APPLICANT: Ferrara, Napoleon
APPLICANT: Filvaroff, Ellen
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Beoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC7
CURRENT APPLICATION NUMBER: US/09/978,189
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/074450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

```

PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084598
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085689
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697
PRIOR APPLICATION NUMBER: 60/085697

Query Match 4.3%; Score 80.5; DB 9; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Gaps 5;

QY 197 DLHDLGHVFWLLHP-----SFSEFFINMGRLF---TKVI-EKVOALPSKKQ 240
DB 464 DAIVDMIGFFDFILEKEPDLDDVVDGYEISDSFFQNNMLNLYNFSAKVMADQURKPPSRDQ 523
QY 241 RIQTLQNLIAIVRCFWFWFVSGLI-----ENHEGRKAYGAVLSSPQELGHA 288
DB 524 WSMTPQT-----VNAYVLTPTKEIVFPAGILQAPFYARNHPKALNFGGIVVMGHELTHA 578
QY 289 FIDNVR 294
DB 579 FDDQGR 584

RESULT 34
US-10-052-586-420
; Sequence 420, Application US/10052586

```

```

Patent No. US20020127584A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Collin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Genlin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C1
CURRENT APPLICATION NUMBER: US/10/052,586
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 22

PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04

PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088722
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088740
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088811
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088825
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088863
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089090
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089908

Query Match 4.3%; Score 80.5; DB 12; Length 736;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 30; Conservative 22; Mismatches 41; Indels 33; Caps 5;

QY 137 DLHDLGHVPLHLH-----SFSEFIMNGRF--TKVI-EKQVALPSKQ 240
DB 444 DAIIDMTGFDFILFKELDDYDGYEISBDSFFONMLNINFSAKYMAOQJRKPPSRQ 523
QY 241 RIOTLQSNLAIYRCFTWVESGLI-----ENHGRKAYGAVLISPOELGHA 288
DB 524 WSMTPQ-----VNAYYLPKNEIVFPAGLLOAPFYARNHPKALNFGIGVWGHETHA 578
QY 289 FIDNVR 294
DB 579 FDDQGR 584

RESULT 35
US-09-888-615-80

```
; Sequence 80, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 765
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-888-615-80

Query Match      4.3%; Score 80.5; DB 10; Length 765;
Best Local Similarity 23.8%; Pred. No. 11; Indels 33; Gaps 5;
Matches 30; Conservative 22; Mismatches 41;

QY 197 DLHDLGHVPLHLP-----SFSEFFINMGRLP---TKVI-EKVQALPSKKO 240
DB 493 DAIDYMGIFDPFILEPKELDDVDYGEI SEDSFQNMNLNLYNFAKVNADQLRKPFRDQ 552

QY 241 RIQTLSNLIARVCFWFTVESGLI-----ENHGRKAYGAVLISSPQELGHA 288
DB 553 WSMTPQT-----VNAYLPTKNEIVFPAGILOQPFYARNHPKALNFGGIGVWVGHELTHA 607

QY 289 FIDNVR 294
DB 608 FDDQGR 613

RESULT 36
US-09-815-242-13304
; Sequence 13304, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11609
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-815-242-11609

Query Match      4.2%; Score 80; DB 10; Length 897;
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 24

Best Local Similarity 19.5%; Pred. No. 16;
Matches 59; Conservative 54; Mismatches 104; Indels 86; Gaps 14;

```
OY 27 PFONSOSLOASTPYRYRILOKENEKQALARK-----63
Db 54 FYKDKK-----PRIVALSQTKRAELEYKONKRDAPKELDIPALEMLK 107
OY 64 ---CISLEFFKMLLFVHLISLKNQREGSTMAVYTPFRNLMWRLSSRSLMK 119
Db 108 MGLTVEVGGFEADVIAISLATLSPKRIYSKXD-----FK-----QLSDKALPD 156
OY 120 SYCPREFLDYLAFAGLL--SDPLDQAVIKFEELETHPSYVSGFVAHQYLSLQDRYF 178
Db 157 GKTEFLAKDCVEKYGILPSQFTDYOGIVG---DSDVYKGVKG--IGSKAKELLO--RLG 210
OY 179 PIASVMTLDKDNFSLTPLDHLGHVFWLLHPSFSEFFINMGRLETKVIEKVOALPSK 238
Db 211 SLEKTYENLDLAKNLSLPMYQAL-----IQDKGSAFLSK 245
OY 239 KQRTQOSNLIA---IVRCFWFTVESGLIENHGRKAYGAVLISSPOLGHR--PIDNV 293
Db 246 E--LAIHERGCIKEPFLSCA--PSENPLIKIKDELKYG--FISTLDLNSPFIYENV 300
OY 294 RVL 296
Db 301 PTL 303

RESULT 38
US-09-731-231A-2
; Sequence 2, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THERMOP
; FILE REFERENCE: CL001007
; CURRENT APPLICATION NUMBER: US/09/731,231A
; NUMBER OF SEQ. ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Human
US-09-731-231A-2
```

Query Match 4.2%; Score 79; DB 10; Length 645;
Best Local Similarity 19.7%; Pred. No. 13;
Matches 55; Conservative 47; Mismatches 107; Indels 70; Gaps 12;

```
OY 3 YCE-RTLDPYIKALKLOSLSFFQN---SOSLQAVSTPYRYRILOKENEKQA 58
Db 110 YCEGRDLDDK-----IOEYKQAGKIFPENQIIEWFIQLLGVDYMERRLIHRDLKSNV 164
OY 59 LARHKCISLEFFKMLLFVHLISLKNQREGSTMAVYTPFR--NRNLMWRLSSRPS 116
Db 165 FLKGNLKLKIDF-----GVSRILMGSCDLATLTITGPHNMSPEALKHGQYTKSD 214
OY 117 LMSYCPREFLDYLAFAGLLSDPLDQAVIKFEELETHPSYVSGFVAHQYLSLQDR 176
Db 215 IWSLACTLYEMCCNHNFAFSNPL--STVLKIVESDT-----PSLPER 255
OY 177 Y-FPIASVMTLDKDNFSLTPLDHLGHVFWL-----LHPSFSEFFINMGRLE--- 224
Db 256 YPKELNAINMESMLNKNPSLRPSAIEL--KIPYDEOQLQNLNMCYSEMTLEDKNDLCKQE 313
OY 225 -----FTKVIEKVOAL--PSKQRTQOTLOS 247
Db 314 AAHINAMQRIHLQTLRALISEVQKMTPRERMRILKQA 352
```

```
RESULT 39
US-09-992-598-84
; Sequence 84, Application US/09992598
; Patent No. US20020160384A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerlitsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavitt, Ivar J.
; APPLICANT: Napiet, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730PIC20
; CURRENT APPLICATION NUMBER: US/09/992,598
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
```

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

```

PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088555
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23

PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFENRNL-----WYRLL-SRFSRLKSYC-----PRFFLDYLEAFGLSD 138
DB 138 TAFPGKLYNEYSYVPGWKWVLLKNSRFYNY-TLCRNGVKKHSKDY--LTD 194
QY 139 FLHQAVIKFELETHFSYP-----VSGVAPH-----QYLSLLQDRYFFPIASVVR 185
DB 195 LITNDVS-SFF--RTSKMYPHPRVLMVISHAAPHGPDSDAPQY-----SRLFFNAS-QH 245
QY 186 TLDKDNESLTPDLIHLLCHVWLLHPSFSEFFINMGLFTKVIKVOALPSKKORIOTL 245
DB 246 ITPSYNAPND-----KEWIMRYGPMKPHME--FTNMLQ-----RKRLOTL 287

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 26

```
OY      246  OSNLIAVRCFWFTVESGLIENHGRKAYGAVLISSPOLGSHAFIDNVRVLPLELDQIR 305
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      288  MSVDDSMETIYNMLVETGELDN-----TYIVTADHGYNIGOFGLVKGKMPYEPD--IR 340
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
OY      306  LPFNTSTP 313
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB      341  VPFYVRGP 348
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 40
US-09-989-293A-84
/ Sequence 84, Application US/09989293A
/ Patent No. US20020177164A1
/ GENERAL INFORMATION:
/ APPLICANT: Ashkenazi, Avi J.
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Bolstein, David
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerltesen, Mary E.
/ APPLICANT: Goddard, Audrey E.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, J Christopher
/ APPLICANT: Gunney, Austin L.
/ APPLICANT: Kijevin, Ivar J.
/ APPLICANT: Napier, Mary A.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: P2730PLC66
/ CURRENT APPLICATION NUMBER: US/09/989,293A
/ PRIOR FILING DATE: 2001-11-20
/ PRIOR APPLICATION NUMBER: 60/049787
/ PRIOR FILING DATE: 1997-06-16
/ PRIOR APPLICATION NUMBER: 60/062250
/ PRIOR FILING DATE: 1997-10-17
/ PRIOR APPLICATION NUMBER: 60/065186
/ PRIOR FILING DATE: 1997-11-12
/ PRIOR APPLICATION NUMBER: 60/065311
/ PRIOR FILING DATE: 1997-11-13
/ PRIOR APPLICATION NUMBER: 60/066770
/ PRIOR FILING DATE: 1997-11-24
/ PRIOR APPLICATION NUMBER: 60/075945
/ PRIOR FILING DATE: 1998-02-25
/ PRIOR APPLICATION NUMBER: 60/078910
/ PRIOR FILING DATE: 1998-03-20
/ PRIOR APPLICATION NUMBER: 60/083322
/ PRIOR FILING DATE: 1998-04-28
/ PRIOR APPLICATION NUMBER: 60/084600
/ PRIOR FILING DATE: 1998-05-07
/ PRIOR APPLICATION NUMBER: 60/087106
/ PRIOR FILING DATE: 1998-05-28
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/087609
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/087759
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/087827
/ PRIOR FILING DATE: 1998-06-03
/ PRIOR APPLICATION NUMBER: 60/088021
/ PRIOR FILING DATE: 1998-06-04

/ PRIOR APPLICATION NUMBER: 60/088025
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088026
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088028
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088029
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088030
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088033
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088326
/ PRIOR FILING DATE: 1998-06-04
/ PRIOR APPLICATION NUMBER: 60/088167
/ PRIOR FILING DATE: 1998-06-05
/ PRIOR APPLICATION NUMBER: 60/088202
/ PRIOR FILING DATE: 1998-06-05
/ PRIOR APPLICATION NUMBER: 60/088212
/ PRIOR FILING DATE: 1998-06-05
/ PRIOR APPLICATION NUMBER: 60/088217
/ PRIOR FILING DATE: 1998-06-05
/ PRIOR APPLICATION NUMBER: 60/088655
/ PRIOR FILING DATE: 1998-06-09
/ PRIOR APPLICATION NUMBER: 60/088734
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088738
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088742
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088810
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088824
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088826
/ PRIOR FILING DATE: 1998-06-10
/ PRIOR APPLICATION NUMBER: 60/088858
/ PRIOR FILING DATE: 1998-06-11
/ PRIOR APPLICATION NUMBER: 60/088861
/ PRIOR FILING DATE: 1998-06-11
/ PRIOR APPLICATION NUMBER: 60/088876
/ PRIOR FILING DATE: 1998-06-11
/ PRIOR APPLICATION NUMBER: 60/089105
/ PRIOR FILING DATE: 1998-06-12
/ PRIOR APPLICATION NUMBER: 60/089440
/ PRIOR FILING DATE: 1998-06-16
/ PRIOR APPLICATION NUMBER: 60/089512
/ PRIOR FILING DATE: 1998-06-16
/ PRIOR APPLICATION NUMBER: 60/089514
/ PRIOR FILING DATE: 1998-06-16
/ PRIOR APPLICATION NUMBER: 60/089532
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089538
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089598
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089599
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089600
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089653
/ PRIOR FILING DATE: 1998-06-17
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/089807
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/089808
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/089947
/ PRIOR FILING DATE: 1998-06-19
/ PRIOR APPLICATION NUMBER: 60/089948
/ PRIOR FILING DATE: 1998-06-19
/ PRIOR APPLICATION NUMBER: 60/089952
```

PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

99 TPFNRM-----WYLL-SSEPSLWKSVC-----PRFELDYLEAFGLLSD 138
138 TAFEGKLYNEINGSVPPGKWEVGLLNRSFYNY-TLCRNGVKEKGSYKDY--LTD 194

QY 139 FLDHQAVIKFPELETHFSYYP-----VSGFVAPH-----OYLSSLQDRYFFIASVMR 185
DB 195 LIITDSV-SFF--RTSKMYPHRVLVMI SHAAPHGPDSPAQY-----SRLEFFNAS-QH 245
QY 186 TLBKNFSLTPDLIHDLLGHVPLLLHPSSEFFINMGRFLTQVIEKVOALPSKKQRIQTL 245
DB 246 ITSINYAANPD-----KHWIMRYTGPMPKHME--FTNNLQ-----RKRLQTL 287
QY 246 QSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRLVPLPLELDOIIR 305
DB 288 MSVDDSMETIYNMLVETGELDN-----TYIVYTADHGYHIGQGLVKGKSMFYEPD--IR 340
QY 306 LPFNTSTP 313
DB 341 VPFYVRGP 348
RESULT 41
US-09-989-735-84
; Sequence 84, Application US/09989735
; Publication No. US20020193299A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Guiney, Austin J.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1C61
; CURRENT APPLICATION NUMBER: US/09/989,735
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607

[illegible]

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

```

, PRIOR FILING DATE: 1998-07-09
Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFFNRNL-----WYRLL-SSRFSLWKSYC-----PRFFLDYLEAPGLLSD 138
DB 138 TAFTEKVLNTEYNGSVYPPGKEWGLKNSRFYNY-TLCRNGVKRKHGSDYSKDY--LTD 194
QY 139 FLDOQAVIKPFELTFHSYYP-----VSGFVAPH-----QYLSILODRYFPPIASVMR 185
DB 195 LITNDSV-SFF--RTSKMYPHRPVLVMTISHAAPHGPDSDAPQY-----SRLFPNAS-QH 245
QY 186 TLDKDNESLTDLHDLGHVPLWLLHPSSEFFNMGRLFTKVKIEKVQALPSKKORIOTL 245
DB 246 TTPSVNAPND-----KHWIMEYTGPMKPIHWE--FTNMLQ-----RKELQTL 287
QY 246 QENLTAIVRCFWFTVESGLIENHEGRKAYCAVLISSPQELGHAFIDNVRLPLELDQIIR 305
DB 288 MSVDOSMETIINMLVETGELDN-----TIVYVADHGHYHQFGLVKGKMPYFED--IR 340
QY 306 LPFNTSTP 313
DB 341 VPFYVRGP 348

RESULT 42
US-09-990-444-84
, Sequence 84, Application US/09990444
, Publication No. US20020193300A1
, GENERAL INFORMATION:
, APPLICANT: Ashkenazi, Avi J.
, APPLICANT: Baker, Kevin P.
, APPLICANT: Rotstein, David
, APPLICANT: Desnoyers, Luc
, APPLICANT: Eaton, Dan L.
, APPLICANT: Ferrara, Napoleone
, APPLICANT: Fong, Sherman
, APPLICANT: Gerber, Hanspeter
, APPLICANT: Gerritsen, Mary E.
, APPLICANT: Goddard, Audrey
, APPLICANT: Godowski, Paul J.
, APPLICANT: Grimaldi, J. Christopher
, APPLICANT: Gurney, Austin L.
, APPLICANT: Kijavir, Ivar J.
, APPLICANT: Napier, Mary A.
, APPLICANT: Pan, James
, APPLICANT: Paoni, Nicholas F.
, APPLICANT: Roy, Margaret Ann
, APPLICANT: Stewart, Timothy A.
, APPLICANT: Tumas, Daniel
, APPLICANT: Watanabe, Colin K.
, APPLICANT: Williams, P. Mickey
, APPLICANT: Wood, William I.
, APPLICANT: Zhang, Zemin
, TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
, FILE REFERENCE: P2730P1C19
, CURRENT APPLICATION NUMBER: US/09/990,444
, CURRENT FILING DATE: 2001-11-14
, PRIOR APPLICATION NUMBER: 60/049787
, PRIOR FILING DATE: 1997-06-16
, PRIOR APPLICATION NUMBER: 60/062250
, PRIOR FILING DATE: 1997-10-17
, PRIOR APPLICATION NUMBER: 60/065186
, PRIOR FILING DATE: 1997-11-12
, PRIOR APPLICATION NUMBER: 60/065311
, PRIOR FILING DATE: 1997-11-13
, PRIOR APPLICATION NUMBER: 60/066770
, PRIOR FILING DATE: 1997-11-24
, PRIOR APPLICATION NUMBER: 60/075945
, PRIOR FILING DATE: 1998-02-25
, PRIOR APPLICATION NUMBER: 60/078910
, PRIOR FILING DATE: 1998-03-20
, PRIOR APPLICATION NUMBER: 60/083322
, PRIOR FILING DATE: 1998-04-28
, PRIOR APPLICATION NUMBER: 60/084600
, PRIOR FILING DATE: 1998-05-07
, PRIOR APPLICATION NUMBER: 60/087106
, PRIOR FILING DATE: 1998-05-28
, PRIOR APPLICATION NUMBER: 60/087607
, PRIOR FILING DATE: 1998-06-02
, PRIOR APPLICATION NUMBER: 60/087609
, PRIOR FILING DATE: 1998-06-02
, PRIOR APPLICATION NUMBER: 60/087759
, PRIOR FILING DATE: 1998-06-02
, PRIOR APPLICATION NUMBER: 60/087827
, PRIOR FILING DATE: 1998-06-03
, PRIOR APPLICATION NUMBER: 60/088021
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088025
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088026
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088028
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088029
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088030
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088033
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088326
, PRIOR FILING DATE: 1998-06-04
, PRIOR APPLICATION NUMBER: 60/088167
, PRIOR FILING DATE: 1998-06-05
, PRIOR APPLICATION NUMBER: 60/088202
, PRIOR FILING DATE: 1998-06-05
, PRIOR APPLICATION NUMBER: 60/088212
, PRIOR FILING DATE: 1998-06-05
, PRIOR APPLICATION NUMBER: 60/088217
, PRIOR FILING DATE: 1998-06-05
, PRIOR APPLICATION NUMBER: 60/088655
, PRIOR FILING DATE: 1998-06-09
, PRIOR APPLICATION NUMBER: 60/088734
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088738
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088742
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088810
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088824
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088826
, PRIOR FILING DATE: 1998-06-10
, PRIOR APPLICATION NUMBER: 60/088858
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/088861
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/088876
, PRIOR FILING DATE: 1998-06-11
, PRIOR APPLICATION NUMBER: 60/089105
, PRIOR FILING DATE: 1998-06-12
, PRIOR APPLICATION NUMBER: 60/089440
, PRIOR FILING DATE: 1998-06-16
, PRIOR APPLICATION NUMBER: 60/089512
, PRIOR FILING DATE: 1998-06-16
, PRIOR APPLICATION NUMBER: 60/089514
, PRIOR FILING DATE: 1998-06-16
, PRIOR APPLICATION NUMBER: 60/089532
, PRIOR FILING DATE: 1998-06-17
, PRIOR APPLICATION NUMBER: 60/089538
, PRIOR FILING DATE: 1998-06-17
, PRIOR APPLICATION NUMBER: 60/089598
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 30

PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02

PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRLN-----WYRL--SSRFLKSYC-----RPFLLYLEAFGLSD 138
DB 138 TAFKRYLNEVNGSYVPGKXKVGILKNSRYNY-TLCRNGVEXKXGSDYKDY--LTD 194
QY 139 FLDHQAVLKPPLETHSYTP-----VSGFVAPH-----QYLSLDQRYFPISVWR 185
DB 195 LITNDV-SFP--RTSKRMTPHRYLVNISHAARPGSDAPQY-----SRLFPNMS-QH 245
QY 186 TLIDKONSLTPDLIDHILGHVPLMLHPSFEFFINMGRLPFKYIEKVALPSKKQRIQT 245
DB 246 ITPSYNAFNPD-----KHWIMRYTQPMKPIHME--FTNMLO-----RKSLQTL 287
QY 246 QSNLIATVRCFWFTVESGLIENHGRKAYGAVLISSPOLGHAFIDNVRLPLEDQIR 305
DB 288 MSVDMSMETTYNMLVETGELDN-----TYIVTADHGHIQFGLVKGKMPYEPD--IR 340
QY 306 LPFNSTSP 313
DB 341 VPFYVRGP 348

RESULT 43
US-09-989-730-84

Sequence 84, Application US/09989730
Publication NO. US20020197674A1

GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Deenoyers, Luc
APPLICANT: Baton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Goddard, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Par, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PIC69
CURRENT APPLICATION NUMBER: US/09/989,730
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250

1	PRIOR APPLICATION NUMBER: 60/089512
2	PRIOR FILING DATE: 1998-06-16
3	PRIOR APPLICATION NUMBER: 60/089514
4	PRIOR FILING DATE: 1998-06-16
5	PRIOR APPLICATION NUMBER: 60/089532
6	PRIOR FILING DATE: 1998-06-17
7	PRIOR APPLICATION NUMBER: 60/089538
8	PRIOR FILING DATE: 1998-06-17
9	PRIOR APPLICATION NUMBER: 60/089598
10	PRIOR FILING DATE: 1998-06-17
11	PRIOR APPLICATION NUMBER: 60/089599
12	PRIOR FILING DATE: 1998-06-17
13	PRIOR APPLICATION NUMBER: 60/089600
14	PRIOR FILING DATE: 1998-06-17
15	PRIOR APPLICATION NUMBER: 60/089653
16	PRIOR FILING DATE: 1998-06-17
17	PRIOR APPLICATION NUMBER: 60/089801
18	PRIOR FILING DATE: 1998-06-18
19	PRIOR APPLICATION NUMBER: 60/089907
20	PRIOR FILING DATE: 1998-06-18
21	PRIOR APPLICATION NUMBER: 60/089908
22	PRIOR FILING DATE: 1998-06-18
23	PRIOR APPLICATION NUMBER: 60/089947
24	PRIOR FILING DATE: 1998-06-19
25	PRIOR APPLICATION NUMBER: 60/089948
26	PRIOR FILING DATE: 1998-06-19
27	PRIOR APPLICATION NUMBER: 60/089952
28	PRIOR FILING DATE: 1998-06-19
29	PRIOR APPLICATION NUMBER: 60/090246
30	PRIOR FILING DATE: 1998-06-22
31	PRIOR APPLICATION NUMBER: 60/090252
32	PRIOR FILING DATE: 1998-06-22
33	PRIOR APPLICATION NUMBER: 60/090254
34	PRIOR FILING DATE: 1998-06-22
35	PRIOR APPLICATION NUMBER: 60/090349
36	PRIOR FILING DATE: 1998-06-23
37	PRIOR APPLICATION NUMBER: 60/090355
38	PRIOR FILING DATE: 1998-06-23
39	PRIOR APPLICATION NUMBER: 60/090429
40	PRIOR FILING DATE: 1998-06-24
41	PRIOR APPLICATION NUMBER: 60/090431
42	PRIOR FILING DATE: 1998-06-24
43	PRIOR APPLICATION NUMBER: 60/090435
44	PRIOR FILING DATE: 1998-06-24
45	PRIOR APPLICATION NUMBER: 60/090444
46	PRIOR FILING DATE: 1998-06-27
47	PRIOR APPLICATION NUMBER: 60/090445
48	PRIOR FILING DATE: 1998-06-24
49	PRIOR APPLICATION NUMBER: 60/090472
50	PRIOR FILING DATE: 1998-06-24
51	PRIOR APPLICATION NUMBER: 60/090535
52	PRIOR FILING DATE: 1998-06-24
53	PRIOR APPLICATION NUMBER: 60/090540
54	PRIOR FILING DATE: 1998-06-24
55	PRIOR APPLICATION NUMBER: 60/090542
56	PRIOR FILING DATE: 1998-06-24
57	PRIOR APPLICATION NUMBER: 60/090557
58	PRIOR FILING DATE: 1998-06-24
59	PRIOR APPLICATION NUMBER: 60/090676
60	PRIOR FILING DATE: 1998-06-25
61	PRIOR APPLICATION NUMBER: 60/090678
62	PRIOR FILING DATE: 1998-06-25
63	PRIOR APPLICATION NUMBER: 60/090690
64	PRIOR FILING DATE: 1998-06-25
65	PRIOR APPLICATION NUMBER: 60/090694
66	PRIOR FILING DATE: 1998-06-25
67	PRIOR APPLICATION NUMBER: 60/090695
68	PRIOR FILING DATE: 1998-06-25
69	PRIOR APPLICATION NUMBER: 60/090696
70	PRIOR FILING DATE: 1998-06-25
71	PRIOR APPLICATION NUMBER: 60/090862
72	PRIOR FILING DATE: 1998-06-26
73	PRIOR APPLICATION NUMBER: 60/090863

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 32

PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFENNL-----WYRL-SSRFSLMKSYC-----PPFLDYLEAFGLLSD 138
DB 138 TAFGKYLNEVNGSYVPPQKEMVGLLNKNSFYNY-FLCNQVKEHGSDYKQY--LTD 194
QY 139 FLDHGVVTFELETHTFSYV-----VSGVAPH-----QYSLLDYRFPFASVWR 185
DB 195 LITNDV-SF--RTSKMTPRPVMTVISAAPHGSDAPQY-----SLRPFVNS-QH 245
QY 186 TLDKDNFSLPDLIHDLGHVWMLHPSEFSEFFINMGRLLTKYIEVQALPSKKQRIQTL 245
DB 246 TTPSYVAPNPD-----KWIWIRYGVPMKPIHME--FTNNLQ-----RRKLOTL 287
QY 246 QSNLIAVRCFWFVESGLIENHEGRKAYGAVLSSPOLGHAFLDNVLEPLDQIIR 305
DB 288 MSYVDSMETVYMLVETGELDN-----YVIYVTDHGHYHIGQGLVKGKSMYEPD--IR 340
QY 306 LPENSTP 313
DB 341 VPFYVGRP 348

RESULT 44
US-09-990-436-84
Sequence 84, Application US/09990436
Publication No. US20020198148A1

GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Geider, Hanspeter
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Sewart, Timothy A.
APPLICANT: Tamas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C14
CURRENT APPLICATION NUMBER: US/09/990,436
CURRENT FILING DATE: 2001-11-14
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25

PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 15;

QY 99 TPFNNRL-----WYRL--SSRFLWKSVC-----PRFFLDYLEAFGLSD 138
DB 138 TAFGKYLNEVNGSVPPGKWKWGLKNSRFNY--TLCRNGVKERKGSYSDY--LTD 194
QY 139 FLDHOAVIKFFLETHESYYP-----VSGFVAPH-----QYLSLLQDRYFPFIASVMR 185
DB 195 LITNSV-SFF--RTSKWYPRPVLWVISHAAPHGSDSAPOY-----SRLFPNAS-QH 245
QY 186 TLDKNFSLTDLHLHLLGHVPLHSPSEFFINNGRLFTKVKVQALPSKKORIOTL 245
DB 246 ITPSYNAPND-----KHWIMRYTGMKPIHME--FTNMLQ-----RKRIOTL 287
QY 246 QSNLIAVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGHAFIDNVRLPLELDOIIR 305
DB 288 MSVDDSMETIYNMLVETGELDN-----TYIVYADHYHIGQFGLVKGKSMPEYFD--IR 340

QY 306 LPFNTSTP 313
DB 341 VPYVRGP 348

RESULT 45
US-09-991-181-84
; Sequence 84, Application US/09991181
; Publication No. US20020197615A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 34

APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tunes, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PIC53
CURRENT APPLICATION NUMBER: US/09/991,181
CURRENT FILING DATE: 2001-11-16
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

```

PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRNL-----WYRLI-SRFSWLKSYC-----PRFFLDYLEAFGLLSD 138
DB 138 TAFCKVLYNEYSVPPCKWEVGLLKNRFRNY-TLCRNGVKHKGSDYSKYD--LTD 194
QY 139 FLHQAVIKFEELTHFSYYP-----VSGFVAPH-----QYLSLLQDRYFFIASVMR 185
DB 195 LITNDSV-SFF--RTSKWYHPRPVLVWISHAAPHGSDAPY-----SRLFFNAS-QH 245
QY 186 TLKDNFSLTPDLIHLGHVWLLHPSFEFFINMGRFTVKVQALPSKKORIOTL 245
DB 246 ITPSYNAPND-----KXIMWRYTGPMPKPHME--FTNMLQ-----RKLQTL 287
QY 246 QSNLIATVRCFTWTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVVLEPLEDQIIR 305
DB 288 MSVDSDMETYNMLNVLGTGLDN-----TIIVTADHGYYHIGQGLVKGKSMYEPD--IR 340
QY 306 LPFNTSTP 313
DB 341 VPFYVRGP 348

RESULT 46
US-09-993-687-84
; Sequence 84, Application US/09993687
; Publication No. US20020198149A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Borstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zeman
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC11
; CURRENT APPLICATION NUMBER: US/09/993.687
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088326
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088167
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088202
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/088212

```


us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

US-09-989-734-84
 Sequence 84, Application US/09989734
 Publication No. US20030003531A1
 GENERAL INFORMATION:
 APPLICANT: Ashkenazi, Avi J.
 APPLICANT: Baker, Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnoyers, Luc
 APPLICANT: Eaton, Dan L.
 APPLICANT: Ferrara, Napoleone
 APPLICANT: Fong, Sherman
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerlitsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Grimaldi, J. Christopher
 APPLICANT: Gurney, Austin L.
 APPLICANT: Kijavini, Ivar J.
 APPLICANT: Napier, Mary A.
 APPLICANT: Pan, James
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Williams, P. Mickey
 APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 FILE REFERENCE: P2730P1C64
 CURRENT APPLICATION NUMBER: US/09/989,734
 PRIOR FILING DATE: 2001-11-19
 PRIOR APPLICATION NUMBER: 60/049787
 PRIOR FILING DATE: 1997-06-16
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/065186
 PRIOR FILING DATE: 1997-11-12
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066770
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/075945
 PRIOR FILING DATE: 1998-02-25
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/084600
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/087106
 PRIOR FILING DATE: 1998-05-28
 PRIOR APPLICATION NUMBER: 60/087607
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087609
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087759
 PRIOR FILING DATE: 1998-06-02
 PRIOR APPLICATION NUMBER: 60/087827
 PRIOR FILING DATE: 1998-06-03
 PRIOR APPLICATION NUMBER: 60/088021
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088025
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088026
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088028
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088029
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088030
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088033
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088326
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088167
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088202
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088212
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088217
 PRIOR FILING DATE: 1998-06-05
 PRIOR APPLICATION NUMBER: 60/088655
 PRIOR FILING DATE: 1998-06-09
 PRIOR APPLICATION NUMBER: 60/088734
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088738
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088742
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088810
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088824
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088826
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088858
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088861
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/088876
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/089105
 PRIOR FILING DATE: 1998-06-12
 PRIOR APPLICATION NUMBER: 60/089440
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089512
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089514
 PRIOR FILING DATE: 1998-06-16
 PRIOR APPLICATION NUMBER: 60/089532
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089538
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089598
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089599
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089600
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089653
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089801
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089907
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089908
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089947
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/089948
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/089952
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/090246
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090252
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090254
 PRIOR FILING DATE: 1998-06-22
 PRIOR APPLICATION NUMBER: 60/090349
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090355

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.rapb

Page 38

PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNNL-----WYLL-SSRFSLMKSYC-----PPFIDYAEFLSLD 138
DB 138 TAFKGYKMEYNGSVYRPMKEMVLLKNSRFYNY-FLCNGYKKGSDYSKDY--LTD 194
QY 139 FLHQAVIKFELETHFSYR-----VSGVAPH-----OYLSLDQRYPPIASVWR 185
DB 195 LITNDVY-SFF--RISKMYRPRVLMYISNAHPGPDASAPOT-----SKLPNNS-OH 245
QY 186 TLKDNFSLTPDLIHDLGHVWLLHPSFSSEFFINMGRLFTKVIKQVALPSKKOIQTL 245
DB 246 ITPSYNAHPD-----KMIKRYGPKPKPIHME--FTNMLQ-----RKLQTL 287
QY 246 QSNLIAIVRCFMFTVESGLIENHGRKAYGAVLLISPOELGHAIFIDNVRVLPLELDQILR 305

DB 288 MSYDSDSHETIYNMLVETGELDN-----TYIYVTDHGVHIGQFGLVYKSKMPYEPD--IR 340
QY 306 LPPHTSR 313
DB 341 VFFYVRGP 348

RESULT 48
US-10-028-072-20
Sequence 20, Application US/10028072
Publication No. US20030004311A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OR INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028, 072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063550
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063733
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063735
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063738
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063755
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064248
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/064809
PRIOR FILING DATE: 1997-11-07
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065846
PRIOR FILING DATE: 1997-11-17
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066453
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066511
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069212
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069278
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069334
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069694
PRIOR FILING DATE: 1997-12-16
PRIOR APPLICATION NUMBER: 60/072320
PRIOR FILING DATE: 1998-01-23
PRIOR APPLICATION NUMBER: 60/073612
PRIOR FILING DATE: 1998-02-04
PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-02-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081695
PRIOR FILING DATE: 1998-04-14
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24

PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085149
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085339
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 19/98-06-11
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090538
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07

Query Match 4.2%; Score 79; DB 9; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TFFFNRL-----WYRL-SRRSLMKSYC-----PRFFLDYLEAFGLSD 138
Db 138 TAFKGLNEYNSYVPPGKMWGLLKNSRFYNY-TLCRNGVKHGSDYSKDY--LTD 194

Fri Jan 10 10:56:11 2003.

us-09-438-185a-1047.rapb

Page 40

```

OY 139 FLHDQAVIKFELETHFSYYP-----VSGFVAPH-----OYLSLDORYPINASVWR 185
OY 195 LITNDV-SFF--RTSKKVPHPRIWVISHAAPHGPEDSAPOV-----SRLEFPMS-QH 245
OY 186 TLDKDNFSLPDILHLLGHVWMLHPSESEFPINMGRLEPKVIEKVALPSKKQRIOTL 245
OY 246 ITTSPYANPDP-----KXIMKYTGPKKPIHME--FTNNMQ-----KSKLOTL 287
OY 246 OSNLIAVRCFMTVESGLIENHEGRKAYGAVLISSPOELGHAFLDNVRLPLELDQIR 305
OY 288 MSYDSDMETTYNNMLVETGELDN-----TYIYVTDHGXYHGOGLVKGSKMPEYFD--IR 340
OY 306 LPENTSTP 313
OY 341 VPFYVRGP 348

RESULT 49
US-09-989-722-84
Sequence 84: Application US/09989722
Patent No. US20020072067A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyer, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertlisen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavini, Ivar J.
APPLICANT: Nadler, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C63
CURRENT APPLICATION NUMBER: US/09/989,722
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
```

PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090676
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090678
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090690
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090694
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090696
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090862
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/090863
PRIOR FILING DATE: 1998-06-26
PRIOR APPLICATION NUMBER: 60/091360
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091478
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091544
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091519
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091626
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091633
PRIOR FILING DATE: 1998-07-02
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/092182
PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 10; Length 867;
Best Local Similarity 23.8%; Pred. NO. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Caps 16;

QY 99 TPFENRNL-----WYRLI-SRFSLSKSYC-----PRFFLDYLEAFGLLSD 138
DB 138 TAFEGKYLNEYNGSVVPPCKWEVGLLKNRFRFNY-TLCRNGVKGKSGSDYSKDY--LTD 194
QY 139 FLDHQAVIKFFFELETHFSYYP-----VSGFVAPH-----QYLSLQDRYFFPIASVMR 185
DB 195 LIITNSV-SFF--RTSKKMYPHRPVLMVISHAAPHGPDSPAFQY-----SRLFPNAS-OH 245
QY 186 TLDDKNFSLTPDLIHDLGHVFWLLHPSFSEFFINMGRFTVKVIEKVAQPSKKRIOTL 245
DB 246 ITSPSYVAPNPD-----KHWIMRVTPGPMKPIHME--FTNMLO-----RKRLQTL 287
QY 246 QSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIR 305
DB 288 MSVDDSNMETIYNMLVETGELDN-----TYIVTTADHGYHIGQFVLKGSMPYEPD--IR 340
QY 306 LPFNTSTP 313
DB 341 VFFVVRGP 348

RESULT 50
US-09-989-723-84
; Sequence 84, Application US/09989723
; Patent No. US20020072092A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Forg, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Williams, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2730P1662
; CURRENT APPLICATION NUMBER: US/09/989, 723
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910

[illegible]

us-09-438-185a-1047.rapb

Fri Jan 10 10:56:11 2003

; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match 4.2%; Score 79; DB 10; Length 867;
Best Local Similarity 23.8%; Pred. No. 19;
Matches 59; Conservative 34; Mismatches 85; Indels 70; Gaps 16;

QY 99 TPFNRNL-----WVRL--SSRFSLWKSVC-----PRFFLDYLEAFGLLSD 138
Db 138 TAFGKYLNEYNGSVYPPGKWKVGLLKNSRFVNY--TLCRNGVKEKHGSDYSKY--LTD 194
QY 139 FLDHOAVIKFSELETHFSYYP-----VSGFVAPH-----QYLSLLQDRYFPPIASVMR 185
Db 195 LITNSV-SFP--RTSKMYFHRPVLVWISHAAPHGPDSDAPQY-----SRLFPNAS-QH 245
QY 186 TLDKONFSLTDPDLHLLGHVPLHHPSEFPFIWGRIFTKVIKQALPCKKORIOTL 245
Db 246 ITFSYNYAPND-----KHWIMRYTGMKFIHME--FTNMLQ-----RKRLQTL 287
QY 246 QSNLIAVRCFWFTVESGLTENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIR 305
Db 288 MSYDDSMETIYNMLVETGELDN-----TYIVYTADGHYHIGQFGLVKGKSWPYEFD--IR 340
QY 306 LPFNTSTP 313
Db 341 VPEYVRGP 348

Search completed: January 9, 2003, 17:13:45
Job time : 52 secs

THIS PAGE BLANK (USPTO)


```

OM protein - protein search, using sw model

Run on: January 9, 2003, 14:46:38 ; Search time 26 Seconds
        (without alignments)
        409,658 Million cell updates/sec

Title: US-09-438-185A-1047
Perfect score: 1989
Sequence: 1 VHYCERTLDPKYLKIALKL.....ESIPLYNQEKYLSGFVLCQ 362

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues      262574
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 150 summaries

Database :
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.      Score  Query Match Length DB ID      Description
-----
1 223 11.8 498 1 US-08-009-075-5      Sequence 5, Appli
2 220 11.6 497 1 US-08-009-075-4      Sequence 4, Appli
3 98.5 5.2 1024 2 US-09-091-117-5      Sequence 5, Appli
4 95 5.0 1226 2 US-08-540-804-12      Sequence 12, Appli
5 95 5.0 1226 3 US-08-218-265-12      Sequence 12, Appli
6 95 5.0 1226 3 US-08-521-872-12      Sequence 12, Appli
7 95 5.0 1226 4 US-08-590-399-12      Sequence 12, Appli
8 92 4.9 4536 4 US-09-180-422B-27      Sequence 27, Appli
9 91.5 4.8 872 3 US-08-766-014-2      Sequence 2, Appli
10 91 4.8 535 3 US-08-369-822C-24      Sequence 24, Appli
11 91 4.8 535 3 US-08-582-776C-19      Sequence 39, Appli
12 91 4.8 535 3 US-08-434-811B-36      Sequence 36, Appli
13 91 4.8 2183 2 US-08-348-891A-7      Sequence 7, Appli
14 91 4.8 2183 2 US-08-905-817-7      Sequence 7, Appli
15 90.5 4.8 615 1 US-08-484-105-6      Sequence 6, Appli
16 90.5 4.8 615 1 US-08-484-105-6      Sequence 6, Appli
17 90 4.8 625 4 US-09-134-001C-4504      Sequence 4504, Ap
18 90 4.8 654 4 US-09-134-001C-3261      Sequence 3261, Ap
19 89 4.7 392 3 US-08-911-853-35      Sequence 35, Appli
20 89 4.7 392 4 US-09-479-409-35      Sequence 35, Appli
21 89 4.7 392 4 US-09-479-453-35      Sequence 35, Appli
22 87.5 4.6 309 3 US-08-996-338-20      Sequence 22, Appli
23 87.5 4.6 521 3 US-08-996-338-20      Sequence 20, Appli
24 87.5 4.6 541 1 US-08-604-333-2      Sequence 2, Appli
25 87.5 4.6 541 3 US-09-110-618-2      Sequence 2, Appli
26 87.5 4.6 541 4 US-09-173-151A-28      Sequence 28, Appli
27 87.5 4.6 541 4 US-09-578-178-2      Sequence 2, Appli
28 86.5 4.6 541 4 US-09-578-178-2      Sequence 2, Appli
29 86.5 4.6 541 4 US-09-578-178-2      Sequence 2, Appli
30 86.5 4.6 541 4 US-09-578-178-2      Sequence 2, Appli
31 86 4.6 541 4 US-09-578-178-2      Sequence 2, Appli
32 84.5 4.5 360 4 US-09-810-347-4      Sequence 4, Appli
33 84.5 4.5 372 4 US-09-810-347-4      Sequence 4, Appli
34 83.5 4.4 1365 6 5194600-4      Patent No. 5194600
35 83 4.4 491 2 US-08-942-819-2      Sequence 2, Appli
36 82 4.3 284 4 US-08-913-159-2      Sequence 2, Appli
37 82 4.3 861 3 US-09-022-875-2      Sequence 2, Appli
38 81.5 4.3 307 4 US-09-810-347-5      Sequence 5, Appli
39 81.5 4.3 608 1 US-08-766-014-3      Sequence 3, Appli
40 81.5 4.3 1708 1 US-08-493-092-2      Sequence 2, Appli
41 81.5 4.3 1708 1 US-08-508-836A-2      Sequence 2, Appli
42 81.5 4.3 3056 2 US-08-508-836A-8      Sequence 8, Appli
43 81.5 4.3 3056 2 US-08-629-001A-3      Sequence 3, Appli
44 81.5 4.3 3056 2 US-08-874-266-2      Sequence 2, Appli
45 81.5 4.3 3056 4 US-08-642-274D-3      Sequence 3, Appli
46 81.5 4.3 3056 4 US-08-952-127-3      Sequence 3, Appli
47 81.5 4.3 3056 4 US-08-952-014C-3      Sequence 3, Appli
48 81.5 4.3 3057 4 US-09-360-416-3      Sequence 3, Appli
49 81 4.3 323 2 US-08-540-804-18      Sequence 18, Appli
50 81 4.3 323 2 US-08-218-265-18      Sequence 18, Appli
51 81 4.3 323 3 US-08-521-872-18      Sequence 18, Appli
52 81 4.3 323 4 US-08-590-399-18      Sequence 18, Appli
53 81 4.3 474 4 US-08-961-083-60      Sequence 60, Appli
54 81 4.3 1118 3 US-08-724-354D-2      Sequence 2, Appli
55 81 4.3 1118 3 US-09-270-984A-2      Sequence 2, Appli
56 80.5 4.3 377 1 US-07-863-169A-1      Sequence 1, Appli
57 80.5 4.3 377 2 US-08-429-964-1      Sequence 1, Appli
58 80.5 4.3 377 3 US-07-935-087-1      Sequence 1, Appli
59 80.5 4.3 377 5 PCT-US93-08062-1      Sequence 1, Appli
60 80.5 4.3 1365 4 US-09-376-330-18      Sequence 18, Appli
61 79.5 4.2 787 1 US-08-574-763-2      Sequence 2, Appli
62 79.5 4.2 981 2 US-08-649-046-2      Sequence 2, Appli
63 79 4.2 207 3 US-08-996-338-23      Sequence 23, Appli
64 79 4.2 674 1 US-08-803-973-7      Sequence 7, Appli
65 79 4.2 674 1 US-08-803-972-7      Sequence 7, Appli
66 79 4.2 707 1 US-08-803-973-12      Sequence 12, Appli
67 79 4.2 707 1 US-08-803-972-12      Sequence 12, Appli
68 79 4.2 1456 1 US-08-803-972-2      Sequence 2, Appli
69 79 4.2 1456 1 US-08-803-972-2      Sequence 2, Appli
70 79 4.2 3418 2 US-08-639-501-2      Sequence 2, Appli
71 79 4.2 3418 2 US-08-603-753D-4      Sequence 4, Appli
72 79 4.2 3418 3 US-09-044-946-2      Sequence 2, Appli
73 79 4.2 3418 3 US-08-755-587-44      Sequence 44, Appli
74 79 4.2 3418 3 US-09-044-908-2      Sequence 2, Appli
75 79 4.2 3418 4 US-09-099-753-4      Sequence 4, Appli
76 79 4.2 3418 4 US-08-986-106-4      Sequence 4, Appli
77 79 4.2 3457 2 US-08-416-603-4      Sequence 4, Appli
78 78.5 4.2 796 4 US-08-737-109-31      Sequence 31, Appli
79 78.5 4.2 3056 4 US-09-360-416-2      Sequence 2, Appli
80 78 4.1 317 4 US-09-327-984A-37      Sequence 37, Appli
81 78 4.1 670 4 US-09-134-001C-2940      Sequence 2940, Ap
82 78 4.1 3066 4 US-08-953-137-12      Sequence 12, Appli
83 77.5 4.1 514 4 US-09-413-814-82      Sequence 82, Appli
84 77.5 4.1 514 4 US-09-413-814-94      Sequence 94, Appli
85 77.5 4.1 577 4 US-09-315-794-32      Sequence 32, Appli
86 77.5 4.1 577 4 US-09-389-341-32      Sequence 32, Appli
87 77.5 4.1 739 4 US-09-134-001C-1586      Sequence 3586, Ap
88 77.5 4.1 905 2 US-09-134-001C-3782      Sequence 3782, Ap
89 77.5 4.1 971 3 US-08-724-354D-22      Sequence 22, Appli
90 77.5 4.1 971 3 US-09-270-984A-22      Sequence 22, Appli
91 77.5 4.1 971 3 US-09-177-431-8      Sequence 8, Appli
92 77.5 4.1 1089 1 US-08-375-300-2      Sequence 2, Appli
93 77.5 4.1 1089 3 US-09-177-431-2      Sequence 2, Appli
94 77.5 4.1 1089 5 PCT-US95-16930-2      Sequence 2, Appli
95 77 4.1 284 2 US-08-424-641B-2      Sequence 2, Appli
96 77 4.1 284 2 US-08-892-980-2      Sequence 2, Appli
97 77 4.1 284 2 US-08-826-439-2      Sequence 2, Appli
98 77 4.1 578 1 US-08-766-014-4      Sequence 4, Appli
99 77 4.1 645 1 US-07-779-172A-3      Sequence 3, Appli
100 77 4.1 749 1 US-08-369-796-8      Sequence 8, Appli

```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 2

101 77 4.1 749 2 US-08-852-091-8 Sequence 8, Appl1
102 77 4.1 749 2 US-08-820-754-8 Sequence 8, Appl1
103 77 4.1 749 3 US-08-856-652-8 Sequence 8, Appl1
104 77 4.1 749 3 US-08-856-869-8 Sequence 8, Appl1
105 77 4.1 749 3 US-09-012-110-7 Sequence 7, Appl1
106 77 4.1 749 3 US-08-848-547-8 Sequence 8, Appl1
107 77 4.1 749 4 US-09-556-273-7 Sequence 8, Appl1
108 77 4.1 749 4 US-08-566-653A-8 Sequence 8, Appl1
109 77 4.1 749 5 PCT-US95-11025-8 Sequence 8, Appl1
110 77 4.1 844 5 US-09-564-805-227 Sequence 227, App
111 77 4.1 1663 2 US-08-793-126-1 Sequence 1, Appl1
112 77 4.1 1663 4 US-09-142-334-22 Sequence 42, Appl1
113 77 4.1 1663 4 US-08-602-359A-41 Sequence 41, Appl1
114 76.5 4.0 311 2 US-08-295-814B-12 Sequence 12, Appl1
115 76.5 4.0 614 1 US-08-295-814B-12 Sequence 8, Appl1
116 76.5 4.0 614 1 US-09-343-361-12 Sequence 12, Appl1
117 76.5 4.0 614 5 PCT-US95-10579-8 Sequence 8, Appl1
118 76.5 4.0 723 4 US-09-134-001C-5060 Sequence 5060, Ap
119 76.5 4.0 1865 1 US-08-588-985-2 Sequence 2, Appl1
120 76.5 4.0 246 4 US-09-134-001C-5292 Sequence 2, Appl1
121 76 4.0 349 1 US-08-118-270-71 Sequence 71, Appl1
122 76 4.0 349 5 PCT-US93-08528-71 Sequence 71, Appl1
123 76 4.0 604 1 US-08-224-657-84 Sequence 84, Appl1
124 76 4.0 604 1 US-08-224-657-95 Sequence 95, Appl1
125 76 4.0 604 1 US-08-354-138-84 Sequence 84, Appl1
126 76 4.0 604 4 US-03-354-138-95 Sequence 95, Appl1
127 76 4.0 604 4 US-08-724-354D-4 Sequence 4, Appl1
128 76 4.0 604 4 US-08-724-354D-4 Sequence 4, Appl1
129 76 4.0 1043 2 US-09-270-984A-4 Sequence 3, Appl1
130 76 4.0 1043 2 US-08-790-519-3 Sequence 3, Appl1
131 76 4.0 2254 2 US-09-134-001C-3703 Sequence 3703, Ap
132 76 4.0 496 4 US-08-757-653-190 Sequence 71, Appl1
133 76 4.0 833 2 US-08-823-516-71 Sequence 132, App
134 76 4.0 833 3 US-08-758-314-132 Sequence 132, App
135 76 4.0 888 4 US-09-134-001C-3032 Sequence 3032, Ap
136 75 4.0 342 4 US-09-330-611-4 Sequence 4, Appl1
137 75 4.0 564 3 US-08-425-843-8 Sequence 8, Appl1
138 75 4.0 1257 1 US-08-049-783-2 Sequence 2, Appl1
139 75 4.0 1257 1 US-08-158-232-6 Sequence 6, Appl1
140 75 4.0 1257 1 US-08-304-626-6 Sequence 6, Appl1
141 75 4.0 1257 1 US-08-316-301A-6 Sequence 6, Appl1
142 75 4.0 1257 2 US-08-611-928-6 Sequence 6, Appl1
143 75 4.0 1257 3 US-09-173-891-6 Sequence 6, Appl1
144 75 4.0 1257 4 US-09-076-137-6 Sequence 6, Appl1
145 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1
146 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1
147 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1
148 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1
149 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1
150 75 4.0 1257 5 PCT-US92-03624-6 Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-08-009-075-5
Sequence 5, Application US/08009075
Patent No. 5300436
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Menek
APPLICANT: WU, Jing
APPLICANT: FRIEDHOFF, David
APPLICANT: FRIEDHOFF, Arnold J.
TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.

ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA
APPLICATION NUMBER: US/08/009, 075
FILING DATE: 19930126
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: TOWNSEND, GUY K.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 498 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-009-075-5
Query March 11.8%; Score 223; DB 1; Length 498;
Best Local Similarity 26.7%; Pred. No. 1.4e-15;
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;
112 SRSRLKWS-----YCPREFLDYEAFLGSLDFLDH-----QAVIKFELETHF 155
Db AEIATKKEYVYTLKELVYTHACREHLEGFOLLERYCYGREDSTI POLEDVSRFLKERTGF 286
Qy 156 SYVPSGVAPHOYLSLDODRYFPPIASVMTLDKDNFSLTPDILHDLGHVPLLPSPS 215
Db 287 QLRVPAALSLARFPLSLAFRVFOCTOYIRHASSPMHSPEDDCHEHLEGHVPLADRTFA 346
Qy 216 EFLFMGRLLTKYIEVKOALPSKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275
Db 347 QFSODI-----LASIGASDEIEKLTST-----VYMFVEFGCKQNGELKAYG 390
Qy 276 AVLLSPQELGHAFIDNVRVLPLELDQIIRLPFNTSTPOETLFSIRF---DELVELTS 331
Db 391 AGLLSYGELHSLSEPEVRAPDDTAAGPYQDDQYQPYVFSSEFNDADKLANVAS 450
Qy 332 KLE 334
Db 451 RIQ 453
RESULT 2
US-08-009-075-4
Sequence 4, Application US/08009075
Patent No. 5300436
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Menek
APPLICANT: WU, Jing
APPLICANT: FRIEDHOFF, David
APPLICANT: FRIEDHOFF, Arnold J.
TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE
TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,075
; FILING DATE: 19930126
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: TOWNSEND, GUY K.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: GOLDSTEIN=1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-009-075-4

Query Match 11.6%; Score 220; DB 1; Length 497;
Best Local Similarity 26.7%; Pred. No. 3e-15;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SRSFSLMKS-----YCPRFPLDYLEAFGLLSDFLDH-----QAVIKPFLETHF 155
DB 226 ABEIATKVVYVTLKGLYATHACGEHLEAFALLERFCYREDNIPQLEDVSRFLKERTGF 285

QY 156 SYTPVSGFVAPHOYLSLDQRYPIASVMTLKDKNFSLTLDLHDLGHVPWLLHPSPS 215
DB 286 QLRPVAGLSARDFLASLAFVPCQYQVIRHASSPMHSPEDCCHELLGHVPMLADRTFA 345

QY 216 EFTINMGLFTKVEKQVALPSKKORIOTLQSNLTAIVRCFWFTVESGLIENHEGKAYG 275
DB 346 QFSQDIG-----LASGASDEIEKLSLTS-----WFTVEFLCKQNGEVKAYG 389

QY 276 AVLSSQFELGHAFIDNVRVLPLELDQIRLPNTSTPQETLFSIRHF-----DELVELTS 331
DB 390 AGLSSYVGLHLLCLSEEPFAIDPAPAAVQYQDQYQSVYVSESFSDAKDLRSYAS 449

QY 332 KLE 334
DB 450 RIQ 452

RESULT 3
US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; VACCINES
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma genitalium
; US-09-091-117-5

Query Match 5.2%; Score 98.5; DB 4; Length 1024;
Best Local Similarity 19.9%; Pred. No. 0.13;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

QY 29 QNSOSLQRAYSTPYRYRIILQENKEK---QALARHKICISILEFFKNLLFVHLLSLSKN 85
DB 564 QQTDSLKNLFSV-----IGDILSETNVNKTITLHAVKNELLSLVETASTLKIKHL----- 613

QY 86 QREGCSTDMAVSTPPFNRLWYLLSSRFSLSKSYCPRFPLDYLEAFGLLSDFLDHQA 145
DB 614 -----NVQYKVLVDKFKLNSFK-----ELNFFPDTKDITPT 647

QY 146 IK--FFLETH-----PSYVPVSGFVAPHOYLSLDQRYPIASVMTLKDKNF 192
DB 648 IKKVLFESENYKTLKKYENEGPGYHMAKFIKVFNSAENTFYSAI-----DKT 698

QY 193 SLTDLIHDLLGHVPWLLHPSPSEFIMNGRLFT-----KVIEKVALPSKKORIOT 244
DB 699 KSIRDLADML--FKLSLESVNDSDSFKINGSFTLKYHGDNLNLLPNYHSLITKNVGYOI 757

QY 245 LOSNLIAIVRCFWFTVESGLIENHEGKAYGAVLISSPOELG----- 286
DB 758 VNVFPHIDARLLTAEIQNTVFSNPK-----FVIKSPVELSKSLFEVWKTIFENSVNQI 810

QY 287 ----HAFIDNVRVLPLELDQIRLPNTSTPQETLFSIRHPD 324
DB 811 LKXYTFKDNLKFFPKADGSGRLSPDLUSKPDQRPVIPPFAVD 852

RESULT 4
US-08-540-804-12
; Sequence 12, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,804
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 4

FILED DATE: 11-OCT-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/521,872
FILING DATE: 21-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH194-03A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-540-804-12

Query Match 5.0%; Score 95; DB 2; Length 1226;
Best Local Similarity 18.8%; Pred. No. 0.4; Indels 204; Gaps 25;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKROSLFONSOLOPASTPYRYRII-----LOKENKEKQALAHKCI 65
DB 13 YILE---KLIFDMTHYNDSOOL-RTWKROISYFLKLGNCYSRLINKE---IFHMLV 65
QY 66 SILEFFKNLFF---VHLLSLKNQREGCSTMAVST-----PFF----- 102
DB 66 EFINKMENFEFLSLHILMIFMNDICQIDNAPVAATITSSQKEPFLVTXITDMLHK 125
QY 103 -----NRLMYRLLSRFSL-----WKS 120
DB 126 YIVSSSKSMINDENYIINDIKNNKIKLNILKLSLILKIPQOSLEVFIFPTSNMEI 185
QY 121 YCPRF-----LDYLEAFGLSDFLDHOAVIKFELETHSYVSGFVAP 166
DB 186 YKPLFEIVSNADTONSDMKKLELISYNESLKNNSIR-----NVIMSANN 236
QY 167 HOYLSLDQRYFPIASV-----RTLDKDN-----FSLTDPDLHDLGHVP 207
DB 237 DFOULTVTCQFPLKSCIQNCIDTQFTKLD-DNPTFDPWPTYVDQNPITMKIIOQL 295
QY 208 WLIHPS--FSEFFINNGRLFTKVIKVOALPSKKQRIQTLSNL--IAIVRCFWPTVESG 263
DB 296 WSIHPSRQPDHYESN-----OLVAKLIL-----RINSTDEDLHEFOLEDAIWSLVFO- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFLD-- 291
DB 344 LAKNFSAQKRVSYVMPSLYRLNLITLTYGIKVPYIRKLISGGLLYLQDSNDKRVHQ 403
QY 292 ---NVRVLPLELDQIIRLPFNSTPOETLFSIRHDELYELTSL----- 333
DB 404 LLINEKISPLMKSQYMMVLRNVMEDVKFELFNPDLVETITQIKRILSNDITNLQLS 463
QY 334 -----EMMLD---OGLLESIPLYNQEKYLSGEVLC 361
DB 464 KTPLSIKIMVAEWYLSHLCGILSV---NRTVLKIKFKIFC 502

RESULT 5
US-08-218-265-12
Sequence 12, Application US/08218265
Patent No. 5922585
GENERAL INFORMATION:
APPLICANT: Young, Richard A.
APPLICANT: Koleske, Anthony J.
APPLICANT: Thompson, Craig M.
TITLE OF INVENTION: No. 5922585el Factors which Modify Gene

TITLE OF INVENTION: Transcription and Methods of Use Thereof
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,265
FILING DATE: 25-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH194-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-218-265-12

Query Match 5.0%; Score 95; DB 2; Length 1226;
Best Local Similarity 18.8%; Pred. No. 0.4;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKROSLFONSOLOPASTPYRYRII-----LOKENKEKQALAHKCI 65
DB 13 YILE---KLIFDMTHYNDSOOL-RTWKROISYFLKLGNCYSRLINKE---IFHMLV 65
QY 66 SILEFFKNLFF---VHLLSLKNQREGCSTMAVST-----PFF----- 102
DB 66 EFINKMENFEFLSLHILMIFMNDICQIDNAPVAATITSSQKEPFLVTXITDMLHK 125
QY 103 -----NRLMYRLLSRFSL-----WKS 120
DB 126 YIVSSSKSMINDENYIINDIKNNKIKLNILKLSLILKIPQOSLEVFIFPTSNMEI 185
QY 121 YCPRF-----LDYLEAFGLSDFLDHOAVIKFELETHSYVSGFVAP 166
DB 186 YKPLFEIVSNADTONSDMKKLELISYNESLKNNSIR-----NVIMSANN 236
QY 167 HOYLSLDQRYFPIASV-----RTLDKDN-----FSLTDPDLHDLGHVP 207
DB 237 DFOULTVTCQFPLKSCIQNCIDTQFTKLD-DNPTFDPWPTYVDQNPITMKIIOQL 295
QY 208 WLIHPS--FSEFFINNGRLFTKVIKVOALPSKKQRIQTLSNL--IAIVRCFWPTVESG 263
DB 296 WSIHPSRQPDHYESN-----OLVAKLIL-----RINSTDEDLHEFOLEDAIWSLVFO- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFLD-- 291
DB 344 LAKNFSAQKRVSYVMPSLYRLNLITLTYGIKVPYIRKLISGGLLYLQDSNDKRVHQ 403
QY 292 ---NVRVLPLELDQIIRLPFNSTPOETLFSIRHDELYELTSL----- 333
DB 404 LLINEKISPLMKSQYMMVLRNVMEDVKFELFNPDLVETITQIKRILSNDITNLQLS 463
QY 334 -----EMMLD---OGLLESIPLYNQEKYLSGEVLC 361
DB 464 KTPLSIKIMVAEWYLSHLCGILSV---NRTVLKIKFKIFC 502

RESULT 6
US-08-521-872-12
; Sequence 12, Application US/08521872
; Patent No. 6015682
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/521,872
; FILING DATE: 31-AUG-1995
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; LENGTH: 1226 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-521-872-12
Query Match 5.0%; Score 95; DB 3; Length 1226;
Best Local Similarity 19.8%; Pred. NO. 0.4;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;
QY 12 YILKALKRQSLFFQNSQSLORAYSTPSYRII-----LQENKEQALAHKCI 65
DB 13 YILE---KLIFDMTHYNDSQL-RTWKQISYFLKLCNCYSLLINK---IFHHLV 65
QY 66 SILEFPKNLFF---VHLLSUKNQRCSTDMVVST-----PFF-----102
DB 66 EFINKMENFEPLSLHILMIFWNDICQIDTNAFVAATITTSQKEPFLVTKTDMLLHK 125
QY 103 -----NRNLWYRLSSRFSL-----WKS 120
DB 126 YIIVSSKSMINDENYIINDIKKKNKIKMLKILSLKILFQSQSLEVFIPFTSNWEI 185
QY 121 YCPRF-----LDYLEAFGLSDFLDQAVIKFELETHFSYYPVSGFVAP 166
DB 186 YKELLFEIVSNADTNQNSDMKKLELISYRNESLKNSSIR-----NVIMSAGNAN 236
QY 167 HQYLSLQDRYEPPLASVM-----RTLDKDN-----FSLTPDLIHLLGHVP 207
DB 237 DFQITVTCQPKPLSCQLNCIDTQTCKLLD-DNTEFDWPTVYDQPLTHMKIQLIL 295
QY 208 WLLHPS--FSEFIMNGRLFTKVEKQVALPSKKQRIQTLQSNL--IAIVRCFWFTVESG 263

DB 296 WSIHPSRQPDHYESN-----QLVAKLLLL-----RINSTDEDLHEFQTEDAINSLVFQ- 343
QY 264 LIENHEGRK-----AYGAV-----LISSP-----QELGHAFID-- 291
DB 344 LAKNFAQRVSVYMMPSLYRLNLNITYGIIKVPYIRKLISSGLLYLODSNDKFVHVQ 403
QY 292 ---NVRVLPLELDQIIRLPENTSTPQETLESIRHFDLVELTSKL----- 333
DB 404 LLINKLISPLMKSQYNWVLNVMEDYVKFYEFNFQDLVEITEIQMKRILSNDIITLQLS 463
QY 334 -----BWLDD-----QGLLESIPLYNQEKVLSGFVLC 361
DB 464 KTLPSIKIMVAEWYLSHLCSGILSV---NRTVLLKIFKIFC 502
RESULT 7
US-08-590-399-12
; Sequence 12, Application US/08590399
; Patent No. 6214588
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,399
; FILING DATE: 26-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/540,804
; FILING DATE: 11-OCT-1995
; APPLICATION NUMBER: US 08/521,872
; FILING DATE: 31-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03A3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1226 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-590-399-12
Query Match 5.0%; Score 95; DB 4; Length 1226;
Best Local Similarity 18.8%; Pred. NO. 0.4;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKALKRQSLFFQNSQSLORAYSTPSYRII-----LQENKEQALAHKCI 65

```

Db 13 YILE--KLFDMKTHVNDSQL-RTWKQOISFLYKLLGNCYSRLINKE---IFHHMLV 65
Qy 66 SILEPKNLLF---VHLISLKNOREGCTDMAVST-----PFF----- 102
Db 66 EFINKNENFEFLPLSHLMLFWNDICOIDNAPVAVITTSQKREFFLVTKITDMLHK 125
Qy 103 -----NENLWYLLSSRFL-----WKS 120
Db 126 YTVSSSKMINDENYIINDIKNNKIKLNTLKLISLLIKIFOQSLEVFIFPYSNWEI 185
Qy 121 YCPFF-----LDYLEAFGLSDFLDHOAVIKFPELETHFSYVSGFVAP 166
Db 186 YKPLFEIVSNADTNQNSDMKKKELLISYRNESLKNNSIR-----NYIMSASNAN 236
Qy 167 HOYLSLDORYEPIASVM-----RTLDKDN-----PSLTPDLJHDLGHVP 207
Db 237 DPOLITVTCQKPKLSCIQNCIDTQFTKLD-DNPTFPMPTVVDQNPJLWHKIIQLIL 295
Qy 208 WLLHPS--PSEFINKGRLEFTVIEKVALPSKKQRIQTOSNL--IAIVRCWFTVESG 263
Db 296 WSIHPSRQPDHYESN-----QVAKLL-----RINSTDEDLHEQIEDALWSLVRC- 343
Qy 264 LIENHGRK-----AYGAV-----LISF-----OELGHAFID-- 291
Db 344 LAKPSKQKVVSYMPSLYRLNLITGLIKVPTIRKLISSGLLYODSNDKRVHQ 403
Qy 292 ---NVRVPLELDQIRLEPNTSTPOETLFSIRHDELVELTSKL----- 333
Db 404 LLINKISPLMKSQYNNVLRNWEYDKVEIFNPDLVETDQIKRRLISNDITNOLS 463
Qy 334 -----EWMLD---QGLLESIPLYNOEKYLSGFEVL 361
Db 464 KTLPSIKIMVAEWYLSHLCGILSSV--NRTVLKIKFKIFC 502

```

```

RESULT 8
US-09-180-422B-27
; Sequence 27, Application US/09180422B
; Patent No. 6444644
; GENERAL INFORMATION:
; APPLICANT: BRUCKDORFER, KARL R
; TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHAYE, P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/180,422B
; FILING DATE: 07-Dec-1998
; CLASSIFICATION: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36663
; REFERENCE/DOCKET NUMBER: 117-268
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 7038164000
; TELEFAX: 7038164100
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-180-422B-27
Query Match 4.9%; Score 92; DB 4; Length 4536;
Best Local Similarity 20.7%; Pred. No. 6.2; Indels 66; Gaps 12;
Matches 61; Conservative 48; Mismatches 118;
3 YCERTLPKYLKIALKRLQSLSLFPNQSLSLQRAVSTPYSYR--IIQKENEKQAL 59
Db 4184 YTBRELCTMPFREVGVLSQVYSKANGSEIL-----FSYFODLVITLFF- 4230
Qy 60 ARHKCISILEPKNLLFVHLISLKNOREGCTDMAVSTPFFNNRLM-----YELLS 113
Db 4231 RKHKLIDIVISMYREL-----LKDSLKBAQEVFKAIQSILKTEVL--RNLDDLQFIFOLLID 4285
Qy 114 RFSLWKSVCPRFADYL--EAFGLSDFLDHOAVIKFPELETHFSYVSGFVAP----- 166
Db 4286 NIKQLKEMKFTYILNVIQEBINTIFNDYIPY--VFKLKENLCLNHHKFEFIQNELQEA 4343
Qy 167 -----HOYLSLDORYEPIASVMRT-----LDKDNFSLTPDLI----- 199
Db 4344 SQELQGHQIIMALREYFDPSIVGTYVTELEKIVSLKKNLVALKDFHSEIVSAS 4403
Qy 200 ---HDLGHVPWLLHPSFSFPIINMGRLEFTVIEKVALPSKKQRIQTOSNLIA 251
Db 4404 NFTSOLSSQVEQFLHRNIQRYLSILTPDGKKEKIELSTPAQEI--IKSQALIA 4456

```

```

RESULT 9
US-08-766-014-2
; Sequence 2, Application US/08766014
; Patent No. 5744312
; GENERAL INFORMATION:
; APPLICANT: Mamone, Joseph A.
; APPLICANT: Davis, Maria
; APPLICANT: Sha, Dan
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/766,014
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 60/008,688
; FILING DATE: December 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 469-1600
; TELEFAX: (213) 955-0440
; TELE: 67-3510
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 872 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-766-014-2

Query Match 4.8%; Score 91.5; DB 1; Length 872;

Best Local Similarity 17.5%; Pred. No. 0.57;
 Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;

QY 14 LKIALKRLQSLSFFONSOSLORAYSTPYRILOKEXKQALARKHCISILEFFKN 73
 DB 224 IKENIENKELAINSKRELATIKRIPIDFEEYKVKFNBEK-----LLELFNK 273
 QY 74 LLFVHLLSLSKNQREGCGTDMVSTPFENRLMYRLLSRFLSKSYCPRFLLDYLEAF 133
 DB 274 LEFFSLDNKKE-----SSIEIVD-----NHKVEKNSK-----VDIKELV 309
 QY 134 GLSDFLDHOAVIKFPELETHPSYFV-----SGFVAPHQVLSLLQRIYPIASV 183
 DB 310 TILQD-----NENIAFYELIVEGEIKKIAFSGKDTVIYDVQTE-----DL 351
 QY 184 MRLDKNFSLTPLDHLI-----CHVPMLLHPSFSEFFINMGLFT 226
 DB 352 KEIFEKEDFETHEIKDFLVRLSYKIGIECKSYIDTAVMAYLLNPSYNY--DLDRVLK 409
 QY 227 KVIKQALPS-----KKRIOTLOSNIIA-----IVRCFW----- 257
 DB 410 KYL-KVD-VPSYGEIFGKGRDKKIEIDENILADYICSRVCYLFDLKELMFIEMDM 467
 QY 258 -----FTVSGLIENHGRKAYGAVLISSP-QELGHAFIDNVRVLPLELDQIR--LPPN 309
 DB 468 KXLLLEIEMPLEVILKSMESVSGFTLDKVELKSLKOKIDDRIGEI---LDKIYKAGYQFN 524
 QY 310 TSTPOE-TLFSIRHFDLVELTSKLEWMLDOGLLESIFLYN-----OEKY 353
 DB 525 VNSPKQLSEFLFKXNLNLFVIAKTKTGSTDSVLEQLVPYNDIVSDIIEYRQLTKLKSTY 584
 QY 354 LSGF 357
 DB 585 IDGF 588

US-08-369-822C-24
 Sequence 24 Application US/08369822C
 Patent No. 6013860

GENERAL INFORMATION:

APPLICANT: Lipkin, W. I.
 APPLICANT: Briesse, Thomas
 APPLICANT: Kliche, Stefanie
 APPLICANT: Schneider, Patrick A.
 APPLICANT: Stiltz, Lothar
 APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,
 TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
 TITLE OF INVENTION: System Diseases
 NUMBER OF SEQUENCES: 46

CORRESPONDENCE ADDRESS:
 ADDRESS: Fulbright & Jaworski, L.L.P.
 STREET: 865 South Figueroa Street, 29th Floor
 CITY: Los Angeles
 STATE: California
 COUNTRY: USA
 ZIP: 90017-2571

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
 SOFTWARE: ASCII (DOS) TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/369,822C
 FILING DATE: 06-JAN-1995

CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Churchill, Margaret A. (Ph.D.)
 REGISTRATION NUMBER: 39,944
 REFERENCE/DOCKET NUMBER: 1279-194XX
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 213/892-9200
 TELEFAX: 213/890-4518

INFORMATION FOR SEQ ID NO: 24:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 535 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-369-822C-24

Query Match 4.8%; Score 91; DB 3; Length 535;
 Best Local Similarity 20.6%; Pred. No. 0.31;
 Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNQREGCGTDMVSTPFENRLMYRLLSRFS 116
 DB 1 EGLTHEQVDNWKSFAGVKF-----GCFMPLSLDSD--LTMVLDKALAAALQR 46
 QY 117 LMKSYCPRRFLDYLEAFGLSLDFLDHOAVIKFPELETHPSYFVSGFVAPHQVLSLLQDR 176
 DB 47 EMDSVYKPEFLRYDPPKGTGS-----RRLVDVFLNDSSFDPPYDVIMVVGAY---LHDP 98
 QY 177 YETIASVMRTLDKDNESLTPDLIHLLGHVPMLLHPSFSEFFINMGLFTKVIKVOALP 236
 DB 99 EFNL-----SYSLQEKIK-----TGELFAKTYKURA-- 127
 QY 237 SKKRIOTLOSNIIA-----IVRCFWFTVBSG-----LIENHGRKAY 274
 DB 128 -----COVIAENLISNGIKYFKDNGMAKQDLTKALHTLAVSGVPKDKESHG--- 178
 QY 275 GAVLISSQBELGHAFIDNVRVLP--LELDQIRLPFTNSTPOE----- 315
 DB 179 GPVLKTSRSPVHTSTNRVRAAKGFPGFQVIRQDQTDHPENMEAYETVSAFITDLKK 238
 QY 316 -----TLFSIRHFDLVELTSKLEWMLDOGLL-----ESIPLY-- 348
 DB 239 YCLNWRVETISLFAQR-LNEIYGLSPFFQWLKRLTSVLVSDPHCPDLDADHIFLYKV 297
 QY 349 -NOE---KY-LSGFVFLCQ 362
 DB 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 11

US-08-582-776C-39
 Sequence 39 Application US/08582776C
 Patent No. 6079510

GENERAL INFORMATION:

APPLICANT: Lipkin, W. I.
 APPLICANT: Briesse, Thomas
 APPLICANT: Kliche, Stefanie
 APPLICANT: Schneider, Patrick A.
 APPLICANT: Stiltz, Lothar
 APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,
 TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
 TITLE OF INVENTION: System Diseases
 NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:
 ADDRESS: Fulbright & Jaworski, L.L.P.
 STREET: 865 South Figueroa Street, 29th Floor
 CITY: Los Angeles
 STATE: California
 COUNTRY: USA
 ZIP: 90017-2576

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 8

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS NT
SOFTWARE: ASCII DOS TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/582.776C
FILING DATE: 04-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
FILING DATE: 06-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/434,831
FILING DATE: 04-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Churchill, Margaret A.
REGISTRATION NUMBER: 39,944
REFERENCE/DOCKET NUMBER: 1279-194C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/680-4518
TELEFAX: 213/682-9200
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEITICAL: NO
ANTI-SENSE: NO
US-08-582-776C-39

Query Match
Best Local Similarity 20.6%, Pred. No. 0.31;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFPKLLFVHLISLSKNQREGCSTDMNAVSTPFNNRLMWRLLSSRFS 116
DB 1 EGLTHQCVDNWKSFGVGF-----GCFMPLSLDSD--LTMVLDKXALALQR 46
QY 117 LKSYCPRRFFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVSGFVAPHQYLSLLDR 176
DB 47 EMDSVYPRKFLRDPKGTGS-----RLVDVFLNDSFDPYDVMVVSAGV---LHDP 98
QY 177 YFPIASVMTLKDKNFSLTPLDHLGHVFWMLHPSFSEFFIMGRLETKVIEKVALP 236
DB 99 EFNVL-----STSLQEKIKK-----TGRLEAKMTYKQRA-- 127
QY 237 SKKQRIQLOSLNLI-----IVRCFWFTVESG---LIENHEGRKAY 274
DB 128 -----CQVIAENLISNGIKYFKDNGMAKDQDLTKALHTLAVSGVKDLKESHG--- 178
QY 275 GAVLISPOELGHAFIDNVRLP--LELDQIIRLPFNSTPOE----- 315
DB 179 GPVLKTYSRPHTSTRNVRRAKGFIPPOVIRDQDQDHPENMEAYETVSATITTDLKK 238
QY 316 -----TLFSIRHPDELVELTSKLEWM---LDOGLL-----ESIPLY-- 348
DB 239 YCLNWRYETISLFAQR-LNEIYGLPSFPQWLHKRLKLETSVLVSDPHCPDPLDAHILPKV 297
QY 349 -NOE---KY-LSGFEVLCO 362
DB 298 PNDQIFIKYPMGIGGYCO 316

RESULT 12
US-08-434-831B-36
Sequence 36, Application US/08434831B
GENERAL INFORMATION:
APPLICANT: Lipkin, M. I.
APPLICANT: Briese, Thomas
APPLICANT: Kliche, Stefanie
```

```

APPLICANT: Schneider, Patrick A.
APPLICANT: Stitz, Lothar
APPLICANT: Schiemann, Anette
TITLE OF INVENTION: Borna Disease Viral Sequences,
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous
TITLE OF INVENTION: System Diseases
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski, L.L.P.
STREET: 865 South Figueroa Street, 29th floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017-2571
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0
SOFTWARE: ASCII (DOS) TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,831B
FILING DATE: 04-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/369,822
FILING DATE: 06-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Churchill, Margaret A. (Ph.D.)
REGISTRATION NUMBER: 39,944
REFERENCE/DOCKET NUMBER: 1279-194C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213/680-4518
TELEFAX: 213/682-9200
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 535 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHEITICAL: NO
ANTI-SENSE: NO
US-08-434-831B-36

Query Match
Best Local Similarity 20.6%, Pred. No. 0.31;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFPKLLFVHLISLSKNQREGCSTDMNAVSTPFNNRLMWRLLSSRFS 116
DB 1 EGLTHQCVDNWKSFGVGF-----GCFMPLSLDSD--LTMVLDKXALALQR 46
QY 117 LKSYCPRRFFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVSGFVAPHQYLSLLDR 176
DB 47 EMDSVYPRKFLRDPKGTGS-----RLVDVFLNDSFDPYDVMVVSAGV---LHDP 98
QY 177 YFPIASVMTLKDKNFSLTPLDHLGHVFWMLHPSFSEFFIMGRLETKVIEKVALP 236
DB 99 EFNVL-----STSLQEKIKK-----TGRLEAKMTYKQRA-- 127
QY 237 SKKQRIQLOSLNLI-----IVRCFWFTVESG---LIENHEGRKAY 274
DB 128 -----CQVIAENLISNGIKYFKDNGMAKDQDLTKALHTLAVSGVKDLKESHG--- 178
QY 275 GAVLISPOELGHAFIDNVRLP--LELDQIIRLPFNSTPOE----- 315
DB 179 GPVLKTYSRPHTSTRNVRRAKGFIPPOVIRDQDQDHPENMEAYETVSATITTDLKK 238
QY 316 -----TLFSIRHPDELVELTSKLEWM---LDOGLL-----ESIPLY-- 348
DB 239 YCLNWRYETISLFAQR-LNEIYGLPSFPQWLHKRLKLETSVLVSDPHCPDPLDAHILPKV 297
QY 349 -NOE---KY-LSGFEVLCO 362
DB 298 PNDQIFIKYPMGIGGYCO 316
```


DB 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 13

US-08-348-891A-7

; Sequence 7, Application US/08348891A

; Patent No. 5654136

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Keiko

; APPLICANT: MORI, Takayuki

; APPLICANT: MAKINO, Satoshi

; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,

; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR

; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: YOUNG & THOMPSON

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER: US/08/348,891A

; FILING DATE: 25-NOV-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/848,400

; FILING DATE: 10-MAR-1992

; APPLICATION NUMBER: JP 3-293625

; FILING DATE: 14-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: KP-7501

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2183 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-348-891A-7

Query Match 4.8%; Score 91; DB 1; Length 2183;

Best Local Similarity 20.6%; Pred. No. 2.6;

Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNORCGCTDMAVSTPFFNRLNMYRLSSRFS 116

DB 429 EGLTHRCQVDNWKSFAGVKF-----GCFMPLSLDS--LTMWLDKALAAALQ 474

QY 117 LWSYCFRFFLDYLEAFGLSLDLDHQAVIKFFFELETHFSYYPVSGFVAPHQYLSLQDR 176

DB 475 EMDSVYKPEFLAYDPKGTGS-----RRLVDVFLNDSDFDYDVMYVVSAGY---LHDP 526

QY 177 YPIASVMTLQKDNFTLPTDLIHLLGHVPLLLHFSFEPFINMGRFLTKVIEKVALP 236

DB 527 EFWL-----SYSLKEIKE-----IVRCFWFTVESG-----LIENHEGRKAY 555

QY 237 SKQRTOTLQSNLIA-----IVRCFWFTVESG-----LIENHEGRKAY 274

DB 556 -----CQVTAENLISNGIKGYFKONGMAKDEHDLTKALHTLAVSGVPKDLKESHRG----- 606

QY 275 GAVLISSPOBLGHAFIDNVRVLP--LELDQIIRLPFTSTPQF----- 315

DB 607 GPVLKTVSRSPVHTSTRNVRAAKGFIGFPQVIRQDQTDHPENMEAYETVSFAFITDLKK 666

QY 316 -----TLFSIRHDELVETLSKLEWM---LQGLL-----ESIPLY-- 348

DB 667 YCLNWRVETISLFAQR-LNEIYGLPSFPQMLHKLRETSVLVYSDPHCPDLDLDAHIPLYKV 725

QY 349 -NOE---KY-LSGFEVLQ 362

DB 726 PNDQIFIKYPMGGIEGYCQ 744

RESULT 14

US-08-905-817-7

; Sequence 7, Application US/08905817

; Patent No. 5824777

; GENERAL INFORMATION:

; APPLICANT: SASAKI, Keiko

; APPLICANT: MORI, Takayuki

; APPLICANT: MAKINO, Satoshi

; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,

; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR

; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: YOUNG & THOMPSON

; STREET: 745 South 23rd Street

; CITY: Arlington

; STATE: Virginia

; COUNTRY: USA

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION NUMBER: US/08/905,817

; FILING DATE: 04-AUG-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/348,891

; FILING DATE: 25-NOV-1994

; APPLICATION NUMBER: US 07/848,400

; FILING DATE: 10-MAR-1992

; APPLICATION NUMBER: JP 3-293625

; FILING DATE: 14-OCT-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: 32,925

; REFERENCE/DOCKET NUMBER: KP-7501A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-521-2297

; TELEFAX: 703-685-0573

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2183 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-905-817-7

Query Match 4.8%; Score 91; DB 2; Length 2183;

Best Local Similarity 20.6%; Pred. No. 2.6;

Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARHKCISILEFFKNLLFVHLLSLSKNORCGCTDMAVSTPFFNRLNMYRLSSRFS 116

DB 429 EGLTHRCQVDNWKSFAGVKF-----GCFMPLSLDS--LTMWLDKALAAALQ 474

MOLECULE TYPE: protein

US-08-484-105-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;
Best Local Similarity 20.7%; Pred. No. 0.43;
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

QY 24 LSPFQNSGLORAYSTPYS-----YRILQENKEKQALARKKCSILEFFNNILP 76
DB 336 MSYFQNAFVSV--FIDPNAVVDPLNDDYLKTL-----SRCPFMFVGLIK 369
QY 77 VH-----LTLSSKNQREGS--TMNAVSTP-----FNRNLWRLLSSRPSLWSKSYC 122
DB 370 QHAPADEILSLTNKRGLEEFFEFLVRENDINGHAKFVARFLSEELNTINLELY- 428
QY 123 PRFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVPV-----SGFVAPH-- 167
DB 429 -----HNLLIGLDSYDLDRMSCKEYKORLHFE--PIDTIFQELFTLDRSGLLTQSIF 480
QY 168 -QYLSLQDXYFFIASVMTLDKNF-SLTPDL--IHDLGHPMLLHP-----SFSEF 217
DB 481 PYSKSNIEDNLSEQVLSLDEKENVDTLSGLDKIMAPVIGQLFKLYREANFTINIVDF 540
QY 218 FNNGRLEPTK--VIEKQALPSKQRIQTLQS-----NLIAVRCFWFTVSGLEIENH 268
DB 541 YIARETLPRSELTNFRDPSNTYKLELATPDPAFDKVALILFQMAIFAFENMGLIK-F 599
QY 269 EGKRAYAV 277
DB 600 QSTKSYDLV 608

RESULT 16

US-08-484-106-6

Sequence 6, Application US/08484106

Patent No. 5614618

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,106
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8771
TELEFAX: (415) 494-8771

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

TELEPHONE: (415) 494-8771
TELEFAX: (415) 494-8771

QY 117 LMKYCPREFLDYLEAFGLSDFLDHQAVIKFPELETHFSYVSGFVAPHOYLSLLODR 176
DB 475 EMDSVYKPEFLKDYDPKTKGS-----RLVYVFLNDSFPPYIMVYVSGAY---LNDP 526
QY 177 YFFIASVMTLDKNFSLTPDLIHDLGHPMLLHPSEFFINMGRLLTKYIEKQALP 236
DB 527 EFWL-----SYSLKEKEIKE-----TRLPAKMTYKRA-- 555
QY 237 SKKRIQTLQSNLIA-----IVRCFWFTVESG---LIENHGRKAY 274
DB 556 -----QVIAMNLISNGIGKXYKONGAKDEHDLTKALHTLAVSGVPKDKESHKG--- 606
QY 275 GAVLISSPOLGHAFIDNVVLP--LELDQIIRLPNTSTPOE-----ESIPLY-- 348
DB 607 GPVLKTYRSRSPVHTSTRNVRRAKGFIFGPQYIRQDQDTHDPENMEAYETVSAFITTDLKK 666
QY 316 -----TLFSIRHPELVELTSKLEMM--LDQGLL-----ESTIPLY-- 348
DB 667 YCLNMYETISLFAQR-LNEIYGLPSFQWILHKLSTSVLYVSDPHCPDLDLHILPLKV 725
QY 349 -NOE--KY-LSGFVLCO 362
DB 726 PNDQIFIKYPMGJIBGYCQ 744

RESULT 15

US-08-484-105-6

Sequence 6, Application US/08484105

Patent No. 5589341

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce
APPLICANT: BELL, Stephen P
APPLICANT: KOBAYASHI, Ryuji
APPLICANT: RINE, Jasper
APPLICANT: FOSS, Margit
APPLICANT: MCNALLY, Francis J
APPLICANT: LAURENSEN, Patricia
APPLICANT: HERSKOWITZ, Ira
APPLICANT: GAVIN, Kimberly
TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,105
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8771
TELEFAX: (415) 494-8771

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

TELEPHONE: (415) 494-8771
TELEFAX: (415) 494-8771

INFORMATION FOR SEQ ID NO: 6;
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-494-106-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;
Best Local Similarity 20.7%; Pred. No. 0.43;
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

QY 24 LSLFQNSQSLQRAYSTPYS-----YVRIILQENKEKQALARKHCISILEFFKNLIF 76
DB 326 MSYFQNAFVS---FIDFVNVDLNDYLLKIL-----SRCTPFMPFVEGLIK 369
QY 77 VH-----LLSLKNGREGCS---TDMVVVSTP-----FFNRNLWYLLSSRSFLWKSVC 122
DB 370 QHAPADEILSLTNKNGLEFFVEFLVRENDINGHAKFVARFLEEBELNITNENLIY- 428
QY 123 PRFFLDYLEAFGLSLDFLDHQAIVKFELETHESYVP-----SQFVAPH-- 167
DB 429 -----HNLGLKLSYLDWSACKKEYKDLHFE--PIDTIFQELFTLDNRSGLLTQSTF 480
QY 168 -OYLSLLODRYFPPIASVMTLDKDNF-SLTPDL-----IHDLLGHVPMLLHP-----SFSEF 217
DB 481 PSYKSNIEDNLLSWEQVLPSEDKENTIDLSGDLKIMAPVLGQDFKLYREANMTINIDF 540
QY 218 FIMGRLFTK--VIEKQALPSKKQRIQTLQS-----NLIALVRCFWFTVESGLIENH 268
DB 541 YIARETLEKEILFIRKDPSTKULEAETPDADFVKVALLFMQAFAPENNGLIK-F 599
QY 269 EGRKAYGAV 277
DB 600 QSTKSYDLV 608

RESULT 17

US-09-134-001C-4504
Sequence 4504, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1997-11-08

PRIOR FILING DATE: 1997-11-08

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4504

LENGTH: 625

TYPE: PR

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4504

Query Match 4.8%; Score 90; DB 4; Length 625;

Best Local Similarity 20.5%; Pred. No. 0.5;

Matches 91; Conservative 71; Mismatches 172; Indels 110; Gaps 23;

11 KYILKIALKRLQSLGSLFFQNSQSLQRAYSTPYSYVRIILQK-----SNKEKQA 58

132 KFDIYLSIKKKQGI-MIDATTSISNAVLHINQITDDDFKVENLILQELPQAHKRIKQI 190

59 LARH---KCSILEFFKNLIFVHLLSKN---QREGCSTD-----MAVSTPFNNLW 107

191 ISKHNDHQLSTSENKIQQLLVHLLIKHSQPEEDWSTDTESLTIKCKIKDINETLG 250

108 YRLSSRSFLWKSVCPRFFLDY-LEAFGL-----LSDFLDH-----QAVI 146

DB 251 YQLNNK-----TSECFSFISYHFNKFDGLIQQLFIQSVDIRLIELMEGHIGFFPSQDTI 305
QY 147 KFELETHES--YYPVSGFVAPHOYLSLLQDRYFP-----IASVMRTLDKD-NFSLTPLD 198
DB 306 LKDNMNVHFSRTYLRMLSMVYLNPLTSQIKRLYPFVENTLYDSIRQLSQDTNIQLSEDE 365
QY 199 IHDLLGHVPMLLHPSPSEF-----FINMGRFLTQVIEKQALPSKKQRIQTLQSNLIA 251
DB 366 IAFETHFOSSIERKSHHVVITACYGLG-ISTLLAEKIKQLNHAIQIVDTLK---LE 421
QY 252 IVRCFWFTVESGLIENHGRKAYGAVLISSPQ-ELGHAFIN-----VRVL--PLE 299
DB 422 DINNYHEGIDDLITTHDFDTSQ---LLOIPKIVQSPFLSPSDEDAKKIEFFVKDMQNPLS 478
QY 300 LDQIIRLPFWTSTPQSTLPSIRHFDDELVELTSKLEWMLD-----OCLLES----- 344
DB 479 KKDILS---KIQLSVESNEFMHNSHILPIPEKSEKILDYHATLDGYIESAIDREKQSS 535
QY 345 -----IPLYNQEKYLSGFEVL 360
DB 536 TYIGKIALPHGNPEKVLKSHMII 559

RESULT 18

US-09-134-001C-3261
Sequence 3261, Application US/09134001C
Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: GTC-007

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 3261

LENGTH: 654

TYPE: PR

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3261

Query Match 4.8%; Score 90; DB 4; Length 654;

Best Local Similarity 20.1%; Pred. No. 0.53;

Matches 83; Conservative 61; Mismatches 152; Indels 116; Gaps 20;

5 ERTLDPKYILKIAL-----KLRQSLSLFFQNSQSLQRAYSTPYSYVRIILQENKEKQAL 59

129 ERLVHPNQILDWILLIEDENLNNEASVOFRNDLNNSAANMTFALSFOHYTMRNERAPLYNL 188

60 ARHKCISILEFFKNLIFVHLLSKNOREGCSITDMVVSTPFNNRLWYLLSSRSFLW 119

189 IKVANDSVLRSEQSVIEGHELPKAKLRGMNAENFWYSSEFCNVHILRAVFIHKSISR 248

120 SYCPFRFFLDYLEAFGLSLDFLDHQAIVK-----FFELETHF-----SYYPVSGFVAP 166

249 -----IQSSNVDDNNVAVKQMPDILIKLEKEFGEDFNQVYHLM--MVHP 291

167 HGYLSLLODRY-----FFIASVMRTLDKDNFSLTPDL-----IHDLLG 204

292 WQIKHQLQSDYRDELBENKILINSHSVPIYAGLSFTLLVPKQDPLSPHIKSLTNVH-ITG 350

205 HYPWLLHPSFEFFINMGRFLTQVIEKQALPSKKQRIQTLQSNLIAIVRCFWFTVESGL 264

351 EIRTL-----SEQTYNGPLVTLQIREIM---SKDEDFSHYQST-----Y 387

265 IENHGRKAYG-----AVLISSPQELG-----HAFIDN--VRVLPLELDQIIRPFNT 310

388 IDENAGIHFYNRNDNEAIQTDREQLGTLFRNNLYQFISNETVPVIFSSL--VATVPYNT 445

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 12

OY 311 STPOETL-----FSIRHDELVELTSKLEMMDO-----GILLESPIYNO 350
Db 446 EADICTLITYONTYQYKNYEBAK-----OMIKDYSKALLGLV--IPLYSK 490

RESULT 19
US-08-911-853-35
Sequence 35, Application US/08911853
Patent No. 6048710
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,853
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,092
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gialster, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-6504
TELEFAX: 650-846-7620
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-911-853-35

Query Match 4.7%; Score 89; DB 3; Length 392;

Best Local Similarity 22.4%; Pred. No. 0.31;

Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;

OY 90 CST-DMAVSTPPFNENLWYLLSSRFSIMKSYCPREFLDYLAFGLLSDPFDHOAVIKP 148
Db 98 CSAGDHVIVRSYVF-----GSTISLFEKYLKRFGEIV--DYPLADLAWQAFAK- 145
OY 149 FELETHPSYVSGFVAPHQYLSLQDRYFPYASVM-----RTLDKNFSLT 195
Db 146 -----PYTKLLFVESPSPNPLAEIVDIGALAEIAHARGALLAVDNCFCCT 188
OY 196 PDLIHDLGHVPLHLPSFSEFFINMGRLEFTYVI-----EKV----- 232
Db 189 PALQOPALGADWVMH-SATKFIQGGRLGGVAVGRAGMEQVGVGLRTAGPTLSPFNA 247
OY 233 -----QALPSKKORIQTOSNLAIYRCFWFTVESGLE-----NHEGRK-----AY 274
Db 248 WMFLKGLFETLRIMQASASALELAK--WLETQPIDRYVYAGLPSHPQHELAKQOQSAF 305
OY 275 GAVL-----ISSPQELGHAFIDNVRVPL-----ELDQIIRLPFNTS-----TPQETLS-IR 321
Db 306 GAVLSFEVKGKKEAMRFIDATRVISITTNIGDTKTTIAHPATTSHGRLSPOERASAGIR 365

OY 322 HFDELVELTSKLEMMLD 338
Db 366 --DNLVAVAGLEDDVD 380

RESULT 20
US-09-479-409-35
Sequence 35, Application US/09479409
Patent No. 6225106
GENERAL INFORMATION:
APPLICANT: Gerritse, Gijbert
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
TITLE OF INVENTION: EXPRESSION LEVELS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/479,409
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/911,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gialster, Debra J
REGISTRATION NUMBER: 33,888
REFERENCE/DOCKET NUMBER: GC361-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-6504
TELEFAX: 650-846-7620
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-479-409-35

Query Match 4.7%; Score 89; DB 4; Length 392;

Best Local Similarity 22.4%; Pred. No. 0.31;

Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;

OY 90 CST-DMAVSTPPFNENLWYLLSSRFSIMKSYCPREFLDYLAFGLLSDPFDHOAVIKP 148
Db 98 CSAGDHVIVRSYVF-----GSTISLFEKYLKRFGEIV--DYPLADLAWQAFAK- 145
OY 149 FELETHPSYVSGFVAPHQYLSLQDRYFPYASVM-----RTLDKNFSLT 195
Db 146 -----PYTKLLFVESPSPNPLAEIVDIGALAEIAHARGALLAVDNCFCCT 188
OY 196 PDLIHDLGHVPLHLPSFSEFFINMGRLEFTYVI-----EKV----- 232
Db 189 PALQOPALGADWVMH-SATKFIQGGRLGGVAVGRAGMEQVGVGLRTAGPTLSPFNA 247
OY 233 -----QALPSKKORIQTOSNLAIYRCFWFTVESGLE-----NHEGRK-----AY 274
Db 248 WMFLKGLFETLRIMQASASALELAK--WLETQPIDRYVYAGLPSHPQHELAKQOQSAF 305
OY 275 GAVL-----ISSPQELGHAFIDNVRVPL-----ELDQIIRLPFNTS-----TPQETLS-IR 321
Db 306 GAVLSFEVKGKKEAMRFIDATRVISITTNIGDTKTTIAHPATTSHGRLSPOERASAGIR 365
OY 322 HFDELVELTSKLEMMLD 338

```
Db 366 --DNLVRVAVGLEDDVD 380
|||: |||: |||:
RESULT 21
US-09-479-453-35
; Sequence 35, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gijbert
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-453-35
Query Match 4.7%; Score 89; DB 4; Length 392;
Best Local Similarity 22.4%; Pred. No. 0.31;
Matches 71; Conservative 40; Mismatches 104; Indels 102; Gaps 16;
QY 90 CST-DNAVSTPFFNNLWYLLSSRSLKSCYCPFFLDYLEAFGLLSDPLDHOAVIKF 148
Db 98 CSAGDHVLSRSVF-----GSTLSLFKRYKRGLEV--DTPPLADUDAWQAQAFK- 145
QY 149 FELETHFSYYPVSGFVAPHQYLSLQDRYPPIASVM-----RTLDKDNFSLT 195
Db 146 -----ENTKLLFVSPSNPLAELVDIGALAEIAHARGALLAVDNCFT 188
QY 196 PDLIHLGLHVPWLLHPSFSFFINMGRLETKVI-----EKV----- 232
Db 189 PALQQLALGADVMWH-SATKFDGQGRGLGGVVGARRAQMEQVVGFLRTAGPTLSPPNA 247
QY 233 ----QALPSKKORIOTLOSNIHAIVRFCFWFTVESGLIE-----NHGRK-----AY 274
Db 248 WMFLKGLGLTRIRMQAQSASALELAR-WLETQGDIDRVVYAGLPSHPOHELAKRQSAF 305
QY 275 GAVL-----ISSPELGHAFIDNVRVLP-----ELDQIIRLPNTS-----TPQETLFS-IR 321
Db 306 GAVLSFVKGKGAARWFDATRVISITNLTGDKTITIAHPATTSHGRLSQAERASAGIR 365
QY 322 HFDELVELTSKLEWMLD 338
|||: |||: |||:
Db 366 --DNLVRVAVGLEDDVD 380
|||: |||: |||:
RESULT 22
US-08-996-338-22
; Sequence 22, Application US/08996338
; Patent No. 6087116
; GENERAL INFORMATION:
; APPLICANT: TORIGOE, Kakuii
; APPLICANT: OKURA, Takanori
; APPLICANT: KURIMOTO, Masashi
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,338
; FILING DATE: 22-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 74,697/1997
; FILING DATE: 12-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 215,488/1997
; FILING DATE: 28-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 291,837/1997
; FILING DATE: 09-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TORIGOE-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 309
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-996-338-22
Query Match 4.6%; Score 87.5; DB 3; Length 309;
Best Local Similarity 21.3%; Pred. No. 0.32;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;
QY 9 DPKYI-----LXIALKLPQSLSLFQNSQSLORAYSTPYVYRIILQKENKEQALAHKC 64
Db 15 EPLYLKHSCSLAHEITTTKSWYKSGQEHVELNPRSSRIAL-----HDC 62
QY 65 ISILEFF-----KNLLFVHLLSLKQREGCSTDMAVVS-----TPSFN-- 103
Db 63 --VLFSEFVPELNDTGSYFFQMKNYTQKWLKLVIRNKHSCFTERQVTSKIVEVKKFFQIT 120
QY 104 -ENLYRLLSSRSLKWSYCPFFLDYLEAFGLLSDPLDHOAVIKFLELTHFSYYPVSG 162
Db 121 CENSYQTLVNSTSLYKN-CKLLLENNK-----NPTIKNAEFEDQ-GYSCVH 168
QY 163 FVAPHQYLSLQDRYPPIASVMRTLDKDNFSLTPDILHDLGHVPWLLHPSFSFFINMG 222
Db 169 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPVLGPKLNHVAVELG 211
QY 223 RLFTKVIKQVQALPSKKORIQTLOSNIHAIVRFCFW-FTVESGLIEN-HEGRKAYGAVLIS 280
|||: |||: |||:
```

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 14

Db 212 K--NVRNCSALNDEVDY-----YMMGEENGSDPNIHEKE---MIM 251
QY 281 SPOELGHAFIDNVRVLPLE 299
Db 252 TPEGKWAH---SKVLRIE 266

RESULT 23
US-08-996-338-20
Sequence 20, Application US/08996338
Patent No. 6087116
GENERAL INFORMATION:
APPLICANT: TORIGOE, Kakuji
APPLICANT: OKURA, Takao
APPLICANT: KURIMOTO, Masashi
TITLE OF INVENTION: POLYPEPTIDES
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE: 22-DEC-1997
APPLICATION NUMBER: JP 74,697/1997
PRIORITY APPLICATION DATA:
FILING DATE: 12-MAR-1997
APPLICATION NUMBER: JP 215,488/1997
PRIORITY APPLICATION DATA:
FILING DATE: 28-JUL-1997
APPLICATION NUMBER: JP 291,837/1997
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L. 25,618
REGISTRATION NUMBER: TORIGOE-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 521
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-996-338-20

Query Match 4.6%; Score 87.5; DB 3; Length 521;
Best Local Similarity 21.3%; Pred. No. 0.7; Indels 95; Gaps 19;
Matches 68; Conservative 50; Mismatches 106;

QY 9 DPKYI-----LKIALKRLSLSLFONSSOLQRAVSTPYRYRIILQENKEKQALARKC 64
Db 15 EPPYLKHCSCSLAHEIETTTKSWYSSGSOEHVELNPRSSRIAL-----HDC 62
QY 65 ISILEFP-----KNLLFVHLISLSKNQREGCSTMAVVS-----TPFN-- 103
Db 63 --VLEFPVPELNDGTSYFQOMKNYQOKKLVNIRNKGSCPTEROVTSKIVEYKFFQIT 120
QY 104 -RLWYRLSSRSLMSKSYCPREFLDYLAEGLLSDFLDQAVIKFPELETHPSYYPVSG 162
Db 121 CENSYYQTLVNSTSLYKN-CKKLLLENNK-----NPTIKQAEFEDQ-GYYSCHV 168

QY 163 FVAPHOYLSLQDRYEPPIASVWRITDKONFSLTPDLIHDLGHVPHILHPSSEFFINNG 222
Db 169 PL--HH-----NKLFLNITK-----TFNIT--IYEDSNIVPVILGPKLNHVAVEIG 211
QY 223 RLFTKVIKQVALPSKKORIOTLQSNLIAIVRCW--FTVESGLIEN-HGCRKAYGAVLIS 280
Db 212 K--NVRNCSALNDEVDY-----YMMGEENGSDPNIHEKE---MIM 251
QY 281 SPOELGHAFIDNVRVLPLE 299
Db 252 TPEGKWAH---SKVLRIE 266

RESULT 24
US-08-604-333-2
Sequence 2, Application US/08604333
Patent No. 576731
GENERAL INFORMATION:
APPLICANT: Parinet, Patricia et al.
TITLE OF INVENTION: Receptor Designated 2P1
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Kathryn A. Anderson, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
FILING DATE: 21-FEB-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Anderson, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2619
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 541 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-604-333-2

Query Match 4.6%; Score 87.5; DB 1; Length 541;
Best Local Similarity 21.3%; Pred. No. 0.75; Indels 95; Gaps 19;
Matches 68; Conservative 50; Mismatches 106;

QY 9 DPKYI-----LKIALKRLSLSLFONSSOLQRAVSTPYRYRIILQENKEKQALARKHC 64
Db 34 EPPYLKHCSCSLAHEIETTTKSWYSSGSOEHVELNPRSSRIAL-----HDC 81
QY 65 ISILEFP-----KNLLFVHLISLSKNQREGCSTMAVVS-----TPFN-- 103
Db 82 --VLEFPVPELNDGTSYFQOMKNYQOKKLVNIRNKGSCPTEROVTSKIVEYKFFQIT 139
QY 104 -RLWYRLSSRSLMSKSYCPREFLDYLAEGLLSDFLDQAVIKFPELETHPSYYPVSG 162
Db 140 CENSYYQTLVNSTSLYKN-CKKLLLENNK-----NPTIKQAEFEDQ-GYYSCHV 187
QY 163 FVAPHOYLSLQDRYEPPIASVWRITDKONFSLTPDLIHDLGHVPHILHPSSEFFINNG 222
Db 168 FL--HH-----NKLFLNITK-----TFNIT--IYEDRSNIVPVILGPKLNHVAVEIG 230

```

Db      231 K---NVRNCSALINEEDVI-----YMFQENGSDPNITHEEKE---MRIM 270
QY      281 SPQELGHAFTDNRVLPLE 299
Db      271 TPEGKWAH-----SKVLRIE 285

```

```

RESULT 26
US-09-173-151A-28
/ Sequence 28 Application US/09173151A
/ Patent No. 632672
/
/ GENERAL INFORMATION:
/
/ APPLICANT: Timans, Jacqueline C.
/ APPLICANT: Debets, Johannes Eduard Maria
/ APPLICANT: Antonius
/ APPLICANT: Sana, Theodore R.
/ APPLICANT: Bazan, J. Fernando
/ APPLICANT: Kastelein, Robert A.
/
/ TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods
/
/ NUMBER OF SEQUENCES: 36
/
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: DNAX Research Institute
/ STREET: 901 California Avenue
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/
/ ZIP: 94304-1104
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30

```

```

1 FILING DATE: 18-MAR-1998
2
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 60/062,066
5 FILING DATE: 15-OCT-1997
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Ching, Edwin P.
8 REGISTRATION NUMBER: 34,090
9 REFERENCE/DOCKET NUMBER: DX0767X
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (650)852-9196
12 TELEFAX: (650)496-1200
13 INFORMATION FOR SEQ ID NO: 28:
14
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 541 amino acids
17 TYPE: amino acid
18 STRANDEDNESS: not relevant
19 TOPOLOGY: linear
20 MOLECULE TYPE: peptide
21
22 US-09-173-151A-28
23
24 Query Match 4.6%; Score 87.5; DB 4; Length 541;
25 Best Local Similarity 21.3%; Pred. No. 0.75;

```


SEQUENCE CHARACTERISTICS:
 LENGTH: 990 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-392-625-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;
 Best Local Similarity 23.1%; Pred. No. 2.4;
 Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSORAYSTPYSY-----RIIL 49
 DB 601 FCPRIIYKNIILPATWKINSEM--FSETENLWNRPATIRKWHIPKDVIIAFGDNRLLL 657
 QY 50 QKENKE-----KQALAHKICISILEFFKNLFLVHLLSKNQREGCSTDMVYSTPFF-- 102
 DB 658 NLLNDKHLIILKELKKGRIIRLESF-----INESNNE-----MLEIVTPLYKK 703
 QY 103 -----NRLWYRLLSRPSLW---KSYCPFFLDYLEAFGL---LSDFLDHQ 143
 DB 704 TSLKEQSFIIIPKRNKHNKDFSIHLSIPKTYQDNFIQDYLLPFITELKVNNFINKF 763
 QY 144 AVIKFEELETHFSYYPVSGFVAPHQYLSLQ--DRYFFIASVMKTLTKDNFSLTPDLI-H 200
 DB 764 FYIKFKEDEDFIK-----LRLREDEYSQIYSFIKNW-KDYCLNLSELYDY 809
 QY 201 DLLGHVP 207
 DB 810 SIVDYVP 816

RESULT 29

US-08-466-961A-20
 Sequence 20, Application US/08466961A
 Patent No. 5843709

GENERAL INFORMATION:
 APPLICANT: Entian, Karl-Dieter
 APPLICANT: Gtz, Friedrich
 APPLICANT: Schnell No. 5843709bert
 APPLICANT: Augustin, Johannes
 APPLICANT: Engelke, German
 APPLICANT: Rosenstein, Ralf
 APPLICANT: Kaletka, Cortina
 APPLICANT: Klein, Cora
 APPLICANT: Wieland, Bernd
 APPLICANT: Kupke, Thomas
 APPLICANT: Kellner, Roland
 TITLE OF INVENTION: Biosynthetic Process for the Preparation of
 Chemical Compounds
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 STREET: 1100 New York Avenue, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,961A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/392,625
 FILING DATE: 22-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/876,791
 FILING DATE: 30-APR-1992

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/784,234
 FILING DATE: 31-OCT-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 0652.0980004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 990 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-466-961A-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;
 Best Local Similarity 23.1%; Pred. No. 2.4;
 Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;
 QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSORAYSTPYSY-----RIIL 49
 DB 601 FCPRIIYKNIILPATWKINSEM--FSETENLWNRPATIRKWHIPKDVIIAFGDNRLLL 657
 QY 50 QKENKE-----KQALAHKICISILEFFKNLFLVHLLSKNQREGCSTDMVYSTPFF-- 102
 DB 658 NLLNDKHLIILKELKKGRIIRLESF-----INESNNE-----MLEIVTPLYKK 703
 QY 103 -----NRLWYRLLSRPSLW---KSYCPFFLDYLEAFGL---LSDFLDHQ 143
 DB 704 TSLKEQSFIIIPKRNKHNKDFSIHLSIPKTYQDNFIQDYLLPFITELKVNNFINKF 763
 QY 144 AVIKFEELETHFSYYPVSGFVAPHQYLSLQ--DRYFFIASVMKTLTKDNFSLTPDLI-H 200
 DB 764 FYIKFKEDEDFIK-----LRLREDEYSQIYSFIKNW-KDYCLNLSELYDY 809
 QY 201 DLLGHVP 207
 DB 810 SIVDYVP 816

RESULT 30

US-08-645-193B-15
 Sequence 15, Application US/08645193B
 Patent No. 5962253

GENERAL INFORMATION:
 APPLICANT: Kupke, Thomas
 APPLICANT: Gotz, Friedrich
 APPLICANT: Kempter, Christoph
 APPLICANT: Jung, Gunther
 TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
 NUMBER OF SEQUENCES: 70
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 STREET: 1100 New York Avenue, Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/645,193B
 FILING DATE: 13-MAY-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 0652.1540000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2500
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 15:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 990 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-645-193B-15

Query Match 4.6%; Score 86.5; DB 2; Length 990;
 Best local similarity 21.7%; Pred. No. 2.4;

Matches 62; Conservative 43; Mismatches 100; Indels 81; Gaps 14;

QY 3 YCERLDPKYLKIAL-KLRQSLSPFONSLORAYST-----PYSYRRI 48
 DB 601 FCPRIYKNIILKPAWKINSEM--FSETEMLNRPATIREKWHIPKQVIAFGDNRL 657
 QY 49 LOKENE-----KQALARKCISILEPEKNIILFVHLISLKNRECSIDMAVSTPPF 102
 DB 658 LNLNDKHIIILKELKHGRIRLESF-----INSENNR-----MELITPLX 703
 QY 103 -----NNLMYRLSSRFSLM--KSYCPREFLDYLEAFGL--LSDFLDH 142
 DB 704 KTSLEQSFIPKNNKHFNNLKDQPSIHLSIPKTYQDNFTQDYLPLTELKANNPLNK 763
 QY 143 QAVIKFELETHFSYVSGVAPHPQVLSLIQ--DRYPIASVMTLKDKNFSLPDLI 199
 DB 764 FVYIKKEDEDFIK-----LRLREDEDSQIYSFIKWM--DQYCLNSELVD 809
 QY 200 HDLGHVPWLLH-----PFSSEFIMNGSLFTKVIKQVALPSK 238
 DB 810 YSIVDYPEVRYVGGPHVIEDIENFMYDSILSINIQSEFIKPK 855

RESULT 31

US-09-134-001C-4322

Sequence 4322, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

FILE REFERENCE: CTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4322

LENGTH: 656

TYPE: PRP

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4322

Query Match 4.6%; Score 86; DB 4; Length 656;
 Best local similarity 22.5%; Pred. No. 1.5;

Matches 51; Conservative 26; Mismatches 78; Indels 70; Gaps 11;

QY 4 CERTLDPK---YILKIALKQSLSPFONSLORAYS-----TPYS---YRI 47
 DB 369 CFTVFNPNPALBEEKVDKLLSPQSEKLRMSFLMKKGLYPIGNLLIHGC 428
 QY 48 ILOKENKQA-----LAHKCISILEPEKNIILFVHLISLKNRECSIDMAVSTP 100
 DB 429 IPVDENGEMESFEIEGERLSRELLDVEHYVRRAPDH-----KESTEDISTDL----- 477
 QY 101 FPNRLMYRLSSRFSLMKSYCPREFLDYLEAFGLSLDLDHQAIVIKFELETHFSYVP 160

DB 478 -----VY-----LMT-----GKYSLSGKQAMTF---EKYFIEDKA 507
 QY 161 SGFAHPHQYSLQDPRYPPIASVMTLKDKNFSLPDLIDHGLGHP 207
 DB 508 SHKEKNPYTYLRKD-----VDMIRKMLKD-FGLNPDGRIINGHP 548

RESULT 32

US-09-810-347-4

Sequence 4, Application US/09810347

Patent No. 6461847

GENERAL INFORMATION:

APPLICANT: YE, Jane et al.

TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

FILE REFERENCE: CLO01169

CURRENT APPLICATION NUMBER: US/09/810,347

PRIOR FILING DATE: 2001-03-19

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 360

TYPE: PRP

ORGANISM: Human

US-09-810-347-4

Query Match 4.5%; Score 84.5; DB 4; Length 360;
 Best local similarity 19.6%; Pred. No. 0.85;

Matches 67; Conservative 51; Mismatches 117; Indels 107; Gaps 14;

QY 45 YRIILOKENKQALARKCISILEPEKNIILFVHLISLKNR---EGCSTDMAVSTP 100
 DB 30 FGRIOQERKN---IRPNIIIVLTDQDVELGSLQVANNKTRKMEHGATPFINAVTTP 85
 QY 101 F-----FNNLMYRLSSRFSLMKSYCPREFLDYLEAFGLSLDLDHQAIV 145
 DB 86 MCCPSRSSMLTGKYVHNHVVYNNENCSPSQMAHEPRFPAYVLNNTGYRTAFG---- 141
 QY 146 IKFPELETHFSYVSGVAPHPQVLSLIQDRYFPPIASVMT--LDKNFSLTPDLIDHL 203
 DB 142 KVLN-ENYNSYIP-PGW---REWLGLIKNSRFYNYVCNGIKKHGPFYADYFTDLI 195
 QY 204 -----GHYPMILHSFSEFIMNR----- 223
 DB 196 TNESINYPKSKMYHRVYVAVISHAAPQEDSAFOFSKLPJNASQATTSPYNAFPM 255
 QY 224 -----LFTKVIKQVALPSKKQRIOTLOSNIILAVRCFWTFVSGSLIE 266
 DB 256 DKHWIMQYTPMLPIHMEFTNLIQ-----RRKDTLMSVDSVRLNMLVETGELE 307
 QY 267 NHEGRKAYGAVLSSPQELGHAFIGDNVAVLPLELDQIIRLPF 308
 DB 308 N-----TYIITYADHGHHIGPGLVKGKSMYDPD--IRVPF 342

RESULT 33

US-09-810-347-2

Sequence 2, Application US/09810347

Patent No. 6461847

GENERAL INFORMATION:

APPLICANT: YE, Jane et al.

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES

FILE REFERENCE: CLO01169

CURRENT APPLICATION NUMBER: US/09/810,347

PRIOR FILING DATE: 2001-03-19

NUMBER OF SEQ ID NOS: 6

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 372

; TYPE: PRT
; ORGANISM: Human
US-09-810-347-2

Query Match 4.5%; Score 84.5; DB 4; Length 372;

Best Local Similarity 19.6%; Pred. No. 0.89; Mismatches 51; Indels 107; Gaps 14;

Matches 67; Conservative 51; Mismatches 117; Indels 107; Gaps 14;

QY 45 YRIILOKNEKEQALAHKICISILBFFKNLLFVHLISLNQR-----ECSTDMAVSTP 100

DB 30 FRGRIQOERKN-----IRPNILIVLDDQVGLSLQWVNRKTRKIMEHGATINAFVTP 85

QY 101 F-----FNRNLWYLLSSRSFLWKS-YCPRFFLDYLEAFGLSLDFLDHOAV 145

DB 86 MCCPSRSMLTGKYVHNHNVTNNENSCSPQAMHEBRTFAVYLANNTGYRTAFTG----- 141

QY 146 IKFPELETHFSYYPVSGFVAPHQYLSLLQDRYFFIASVMRT--LDKDNFSLPDLIHDLL 203

DB 142 -KYLN-EVNGSYTP-PGW---REWLGLIKNSRFNYITVCRNGIKKHEKGFYADYFTDLI 195

QY 204 -----CHVPWLLHPSFGEFFINMGR----- 223

DB 196 TNESINYFKSKMYRHPVPMVISHAAPHGPDSPQSKLYPNASQHIPTSYNAPNM 255

QY 224 -----LTKVIEKVOALPSKQRIOTLOSNIATVRCFWFTVESGLIE 266

DB 256 DKHIMQVTGPMPIHMEFNILQ-----RKKLQTLMSVDDSVRLYNLIVETGELE 307

QY 267 NHEGRKAYGAVLISSPOELGHAFIDNVRVLPLELDQIIRLPF 308

DB 308 N-----YIIIVTDHGVHIGQGLVKGKSNPYDFD--IRVPF 342

RESULT 34

5194600-4

; Patent No. 5194600

; APPLICANT: BUSSEY, HOWARD; BOONE, CHARLES; SOMMER, STEVE S.;

; HILL, KATHRYN; MEADEN, PHILIP

; TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN

; ASSEMBLY AND USE THEREOF

; NUMBER OF SEQUENCES: 4

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/488,316

; FILING DATE: 05-NAR-1990

; SEQ ID NO.4:

; LENGTH: 1365

5194600-4

Query Match 4.4%; Score 83.5; DB 6; Length 1365;

Best Local Similarity 19.0%; Pred. No. 0.83; Mismatches 49; Indels 167; Gaps 19;

Matches 74; Conservative 49; Mismatches 100; Indels 167; Gaps 19;

QY 89 GCSTDMVSTP--FNRNLWYLLSS--RSLWKSVC----- 122

DB 172 GCPT---VIDSDFEFNKLPHEAMNGEGKFRP-IRWSTCSLDGKSVYPLTHPLEITLQ 227

QY 123 -----PRFLD-----YLEAFGLSLDFLDH--- 142

DB 228 NGRSMSSIPQLKILYTPKPEILVGADNKKQLHLEPEELRELKLRVTSLSIEFYQKDK 287

QY 143 -----QAVIKFF-----ELETHFSYYPVSGF----- 163

DB 288 ITATLFTKSIIVNNPPLISKOLIKVSVNKKIITSNEELNSKGFYNNLMGLYINGONWKI 347

QY 164 --VAPHOYLSLLQDRYFFIASVMRTDK-----DNFSLPDLIHDLLGHVPWL----- 209

DB 348 TSLTPNLLTALKTEYQSLLKITNLLQLELPSKICLLDSKFLNKFQSLKQLONLOPIK 407

QY 210 --LH--PSPSEFFINMGRL-----FTKVIKVOAL--PSKQRIOTLOSNIATVRCFWF 258

DB 408 MDLHTIPGSESIVFNDIESDPQYDELVNSVQAFDKSKFGELPEIKQWSEI11---F 463

QY 259 TVESGLIENHEGRKAYG-----AVLISS---POELGHAFIDNVRVLPLELDQIIRLPENTS 311

DB 464 VIDFARLEDESEVKALGLVRAVNVVSQGYPQVRG-----LLPFSSD 505

QY 312 TPQETLFSIRHF---DELVELTSKLEWML 337

DB 506 SDRSVUNKIYELKKNSTDNLTELKSFLETML 535

RESULT 35

US-08-942-819-2

; Sequence 2, Application US/08942819

; Patent No. 5965707

; GENERAL INFORMATION:

; APPLICANT: Tam, See-Ying

; APPLICANT: Teal, Mandy

; APPLICANT: Galli, Stephen J.

; TITLE OF INVENTION: RIN2, A NOVEL INHIBITOR OF

; TITLE OF INVENTION: RAS-MEDIATED SIGNALING

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: MA

; COUNTRY: USA

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows 95

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/942,819

; FILING DATE: 02-OCT-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/058,520

; FILING DATE: 11-SEP-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Granahan, Patricia

; REGISTRATION NUMBER: 32,227

; REFERENCE/DOCKET NUMBER: BIH96-13pa

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 781-861-6240

; TELEFAX: 781-861-9540

; TELEX:

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 491 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; US-08-942-819-2

Query Match 4.4%; Score 83; DB 2; Length 491;

Best Local Similarity 18.4%; Pred No. 2;

Matches 60; Conservative 48; Mismatches 110; Indels 108; Gaps 12;

QY 15 KIALKROSLFLPQNSLORAYSTPYPIRILLOKENEKQALAHKICISILEFFKNL 74

DB 56 ELAERLQREEREPASSQSQGQSUTTSKFE---EKTNEK---TRKTTVKKFF--- 105

QY 75 LEVHLLSLKSNQRCGSDTM--AVSTPFFNRLWYLLSSRFLSKSYCPREFLDYLEA 132

DB 106 -----SASSRAGSKKEIOEAKAPSPSINR-----OTSIEDRTVTKFIDFLKT 148

QY 133 FGLLSDFLDH---QAVIKFFELETHFSYYPVSGFVAPHOYLSLLQDRYFFIASVMRTDK 189

DB 149 F-----HKTGQEVTK--QTKWPLEAMPYKRDLSIEQSECTODFYQNVARMQT--- 195

QY 190 DNFSUTPOLIHDLLGHVPWLLHPSFSEFFINMGRFLTIVIEKVOALPSKKQRIOTLOSNI 249

Fri Jan 10 10:56:11 2003

US-09-438-185a-1047.ral

Page 20

Db 196 -----RGKVP-----PENYEKIM----- 208
Qy 250 IAYRCWFVWESGLIENHKGKAVANL---ISSPOLGHAFIDNVRVLPLELDQITRL 306
Db 209 -----DQLEKHMTRLYKVFCEETIDDEKDLAQKIRALHWVTQMLCV 255
Qy 307 PFNTSTPQETLFSIRHFDLVELTSG 332
Db 256 PVNEEIPVSDMVAVKAITDIIMDSK 281

RESULT 36
US-08-913-159-2
Sequence 2, Application US/08913159
Patent No. 6300109
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Plasmid-derived type II
TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis
NUMBER OF SEQUENCES: 14
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,159
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 0179/95
FILING DATE: 17-FEB-1995
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-159-2

Query Match .4.3%; Score 82; DB 4; Length 284;
Best Local Similarity 26.1%; Pred. No. 1.1; Mismatches 69; Indels 52; Gaps 12;
Matches 53; Conservative 29; Mismatches 69; Indels 52; Gaps 12;
Qy 184 MRLDKNPFSTPDL-----IHDLGHVPLMLHPSFSF---FINMGRLEPTKYIEKQVL 235
Db 1 MNLQKXKILRPFKWTGKQQLPHIQYLMPEKINFPEPFIGGALPFE-----L 53
Qy 236 PSKKORIQLQSNLAIYRCFWFVWES---GLIENH---BGRKAVGVLISSPOLGHAFID 291
Db 54 APOKAVINDENSELINCYROMKNPEQLIELTNHORENSKEYYLDLRSSDRD---GRID 110
Qy 292 NVRVLPLELDQITRLPFNTSTPQETLFSIR-HFDLVELTSGKLEW-----MLDQ 339
Db 111 KMS-----EVERARARI-----MYMLRVDFNGLYKVNKQFNVPYGRYKMPKIVDK 156
Qy 340 GLEESIPLY---NOEKYLSG-FE 358
Db 157 ELIESISEYLNNSIKIMSGDFE 179

RESULT 37
US-09-022-875-2
Sequence 2, Application US/09022875
Patent No. 6071894
GENERAL INFORMATION:
APPLICANT: Wallis, Nicola G.
TITLE OF INVENTION: No. 6071894el Compounds
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert, Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia

STATE: PA
COUNTRY: USA
ZIP: 19103-2793
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,875
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/039,478
FILING DATE: 25-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Falk, Stephen T
REGISTRATION NUMBER: 36,795
REFERENCE/DOCKET NUMBER: GMS0009
TELEPHONE: 215-994-2488
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 861 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-022-875-2

Query Match 4.3%; Score 82; DB 3; Length 861;
Best Local Similarity 22.4%; Pred. No. 6; Mismatches 144; Indels 92; Gaps 22;
Matches 87; Conservative 66; Mismatches 144; Indels 92; Gaps 22;
Qy 11 KYILKIALKIKROSLSLFPQNSOLQAVSTP-----YSYRI--ILQKE-----NREKA 58
Db 488 KFOYSITFKQLYRTDILFQFNDSIKQTYVENMLINAGYQINLOOSITIVYINOSKYI 547
Qy 59 LARHKCISILEFFKNLIFVLLSLIS---KNQRE-GCSTDMVAVSTPFENNMLYRLLSSR 114
Db 548 -----KTIPLQNHIDNTQOHEQLSVIKNEROGATTD-----TFPGINKV--LIPIG 595
Qy 115 FSIKMSYCPREFLDYLEAFGLSDPLDQA--VIKFEF---LETHFSYVSGVAPHOY 169
Db 596 TSPK-----GILA--IDYSSQVIMPYDASILESMNLBSLA-----VEN 634
Qy 170 LSLDQRFPLASVMTLKDNE--SLTPDL---IHDLGHVPLMLHPSFSFIFIMGR 224
Db 635 VTLKQRESMLQEROLTHSNFLKSHDITPULTTIMGMDLYLVSHKD----- 685
Qy 225 FTKVIEKVALP--SKKQRIQLQSNLAIYRCFWFVWES-----SGLIEN---HE 269
Db 686 -MSLIEKQQLVHSPFOSSOYLTLVNTLISLTKQSSNVQIKQOPYVLSLVEIMDLE 744
Qy 270 GRKAVGVLISSPOLGHAFIDNVRVLPLELDQITRLPFNTSTPQETLFSIRHFD-ELVE 328
Db 745 RHHLKRTIVSSSVNLOPHIDSKLILQALFNLINAVKHTSTDTKINLSIRVASYEOIE 804
Qy 329 LTSKLEWMLDQGLIESIPLYNQEYLSGF 357
Db 805 FA-----VIDEG--PGISLEQOKIPEPF 826

RESULT 38
US-09-810-347-5
Sequence 5, Application US/09810347
Patent No. 6461847
GENERAL INFORMATION:
APPLICANT: Ye, Jane et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
TITLE OF INVENTION: THERIOF

FILE REFERENCE: CLO01159
 CURRENT APPLICATION NUMBER: US/09/810,347
 CURRENT FILING DATE: 2001-03-19
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 5
 LENGTH: 307
 TYPE: PRT
 ORGANISM: Human
 US-09-810-347-5

Query Match 4.3%; Score 81.5; DB 4; Length 307;
 Best Local Similarity 19.8%; Pred. No. 1.4;
 Matches 57; Conservative 39; Mismatches 93; Indels 99; Gaps 12;

QY 95 AVVSTPF-----FNRNLWYLLSSRFLWKS-YCPFFLDYLEAFGLLSDF 139
 DB 27 APVTTCPCPSRSMLTGKYVHNHNVYTNNECSPSQAMHEPRTFAVYLNNTGYRTAF 86
 QY 140 LDHOAVIKFPELETHRSYYPVSGFVAPHOYLSLQDRYFPIASVMRT--LDKDNFSLTPD 197
 DB 87 FG-----KYLN-EYNGSYP-PCW---REWLGLKNSRFYNTVCRNGIKEKGFDYAKD 136
 QY 198 LHHDL-----GHVPMLLHPSFSEFFINMGR----- 223
 DB 137 YFTDLITNESINYFMSKMYPRPVMWVISHAAPHGDSAPQFSKLYNASQHITPSY 196
 QY 224 -----LFTKVIKQVALPSKKRIOTLQSNLIAIVRCFWFTV 260
 DB 197 NVAPNMDKHWIMQVTGPMLEPIHMEFTNIIQ-----RKRLQTLMSVDSVERLYNMLV 248
 QY 261 ESSLIEHGRKAYGAVLSSPOLGHAFIDNVRVLPLELDQIIRLPF 308
 DB 249 ETGELEN-----TYIITADGHYHIGQFGLVKGKMPYDFD--IRVPF 289

RESULT 39
 US-08-766-014-3
 Sequence 3, Application US/08766014
 Patent No. 5744312
 GENERAL INFORMATION:
 APPLICANT: Mamone, Joseph A.
 APPLICANT: Davis, Maria
 APPLICANT: Sha, Dan
 TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE
 NUMBER OF SEQUENCES: 28
 CORRESPONDENCE ADDRESS:
 ADDRESS: Lyon & Lyon
 STREET: 633 West Fifth Street
 STREET: Suite 4700
 CITY: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071-2066
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/766,014
 FILING DATE: Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION DATA: including application
 PRIOR APPLICATION DATA: described below:
 APPLICATION NUMBER: 60/008,688
 FILING DATE: December 15, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Warburg, Richard J.
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 223/104

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 608 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-766-014-3

Query Match 4.3%; Score 81.5; DB 1; Length 608;
 Best Local Similarity 18.0%; Pred. No. 4;
 Matches 67; Conservative 57; Mismatches 120; Indels 129; Gaps 18;
 QY 65 ISILEFFKNLLFVHLLSLSKNQREGCTDMAVVSTPFFNRNLWYLLSSRFLWKSICYCPR 124
 DB 1 MKLLEFNKLEFFSLIDNIKKE-----SSIIVD-----NHKVEKWSK----- 38
 QY 125 FFLDYLEAFGLLSDFLDHOAVIKFPELETHRSYYPV-----SGFVAPHQYLSLQ 174
 DB 39 --VDIKELVTLQD-----NRTAFYPLIVEGEIKKIAFSFGKDTVIDVFQ 83
 QY 175 DRYFPIASVMRTDKDNFSLTPDLIHDLL-----GHVPMLLHPSFSEF 217
 DB 84 TE-----DLKEIFEKEDFEFTTHEIKDFLVLSYKGIKSKYIDTAVMAYLLNPSESNY 138
 QY 218 FINMGRFLTQVIEKVOALPS-----KKRIOTLQSNLIA--IVRCF----- 257
 DB 139 --DLDRVLKYL-KVD-VFSYEGIFGKGRDKKIEIDENILADYICSRVYFLDKLKEKL 194
 QY 258 -----FTVESGLIENHEGRKAYGAVLSSPOLGHAFIDNVRVLPLELDQI 303
 DB 195 MNFIEEMDKKLLLEIMPELVLSKMEVSGFTLDEKVLKELSKQIDDRIGEI--LDKI 251
 QY 304 IR---LPFNTSTPQE-TLFSIRHFDLVELTSKLEWMLDQGLLESIPLYN----- 349
 DB 252 YKEAGYQFNVNSPQJSEFLFEKLNLFVTKTKTGYSTDSEVLQVLPYNDIVSDIIEYR 311
 QY 350 -----QEKYLSGF 357
 DB 312 QUTKLKSTYIDGF 324

RESULT 40
 US-08-493-092-2
 Sequence 2, Application US/08493092
 Patent No. 5728807
 GENERAL INFORMATION:
 APPLICANT: Shiloh, Yosef
 APPLICANT: Tagle, Danilo A.
 APPLICANT: Collins, Francis S.
 TITLE OF INVENTION: Ataxia-Telangiectasia Gene
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESS: Reising,
 STREET: P.O. Box 4390
 CITY: Troy
 STATE: Michigan
 COUNTRY: US
 ZIP: 48099
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/493,092
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:

NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-310 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-493-092-2

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 1; Length 1708;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLROSLSLFFONSOSLQRAVSTPYRYRIIL---OKENKKEQ 57
DB 34 PSHVIAKATPAVYISNCHKTKKLSILEILSK-----SPDSYQKILLALICEQAETNN 83
QY 58 ALARHKCISILEFFKNLFFVHLSLSKNQREGCTDMAVVSTPEFRNIMLYRLSSRFSL 117
DB 84 VYKGRILKTYHFLVSL-----KDKSGLGAMA-----FVLADVIYTLIH----- 126
QY 118 WKSYPFRFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYTPVSGFVAP----- 166
DB 127 YINORPSCIMDVLSRFSLCDDLSCOVCTAVYCKDALENHL--HVIVGTLLPLVYEQV 184
QY 167 ---HOYLSLLQDRYFPPIASVMTLKDKNFSLTPDLIHDLGHV----- 206
DB 185 EVQKQVLDL--KYL-----VIDNKNENLYITTKLDLPFDHVFVKDLRTQOKIKYSRG 238
QY 207 PMLHPSEFPFNNMGRLEFTKVIKVOALPSKKORIOTLOSNIJAIVRCEFTVESGLIE 266
DB 239 PFSLEIEINH-FLSVSVVDALPLTRLEGKDLRQLERLHKQWVIDIRASODNPQDGINV 297
QY 267 -----NHEGRK---AYGAVL 278
DB 298 KLVVNLQLSKMAINHTEKEVLEAVGSL 327

RESULT 41
US-08-508-836A-2
Sequence 2, Application US/08508836A
Patent No. 5777093
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
TITLE OF INVENTION: Ataxia-telangiectasia Gene
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reising, Ethington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,836A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-313 (TAU)

TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1708 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-508-836A-2

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 1; Length 1708;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLROSLSLFFONSOSLQRAVSTPYRYRIIL---OKENKKEQ 57
DB 34 PSHVIAKATPAVYISNCHKTKKLSILEILSK-----SPDSYQKILLALICEQAETNN 83
QY 58 ALARHKCISILEFFKNLFFVHLSLSKNQREGCTDMAVVSTPEFRNIMLYRLSSRFSL 117
DB 84 VYKGRILKTYHFLVSL-----KDKSGLGAMA-----FVLADVIYTLIH----- 126
QY 118 WKSYPFRFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYTPVSGFVAP----- 166
DB 127 YINORPSCIMDVLSRFSLCDDLSCOVCTAVYCKDALENHL--HVIVGTLLPLVYEQV 184
QY 167 ---HOYLSLLQDRYFPPIASVMTLKDKNFSLTPDLIHDLGHV----- 206
DB 185 EVQKQVLDL--KYL-----VIDNKNENLYITTKLDLPFDHVFVKDLRTQOKIKYSRG 238
QY 207 PMLHPSEFPFNNMGRLEFTKVIKVOALPSKKORIOTLOSNIJAIVRCEFTVESGLIE 266
DB 239 PFSLEIEINH-FLSVSVVDALPLTRLEGKDLRQLERLHKQWVIDIRASODNPQDGINV 297
QY 267 -----NHEGRK---AYGAVL 278
DB 298 KLVVNLQLSKMAINHTEKEVLEAVGSL 327

RESULT 42
US-08-508-836A-8
Sequence 8, Application US/08508836A
Patent No. 5777093
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
TITLE OF INVENTION: Ataxia-telangiectasia Gene
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Reising, Ethington, Barnard & Perry
STREET: P.O. Box 4390
CITY: Troy
STATE: Michigan
COUNTRY: US
ZIP: 48099
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,836A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: P-313 (TAU)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 689-3500
TELEFAX: (810) 689-4071

INFORMATION FOR SEQ ID NO: 8;
SEQUENCE CHARACTERISTICS:
LENGTH: 3056 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-629-001A-3

Query Match 4.3%; Score 81.5; DB 1; Length 3056;
Best Local Similarity 18.5%; Pred. No. 47;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYIKIAL-----KLRQSLFFQNSQSLQRAYSTPYSYVRIIL---QKENKEQ 57
DB 1382 PSHVIKATPAYISNCHTKLKSLILSK-----SPDSYQKILLAIQEAETNN 1431
QY 58 ALARHKCISILEFFKNLLFVHLLSLSKNORECSTDMVSTPFFNRNLWYLLSRFSL 117
DB 1432 VYKGRILKIYHLFVSLLL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1474
QY 118 WKSVCPRFFLDY-LEAFGLLSDFLDH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166
DB 1475 YINQRPSCIMDVSLRSFSLCCDLLSQVCQTAVTYCKDALENHL--HVIVGTLLIPLVYEQV 1532
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206
DB 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFDPHVFKDLRITQOKIKYSG 1586
QY 207 PHLHPSEFFINMGRLETKVIEKQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266
DB 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRRQLELHKDQVDMRASQDNPQDGINV 1645
QY 267 -----NHEGRK-----AYGAVL 278
DB 1646 KLVNLLQLSKMAINHTEGEVLEAVGSCL 1675

RESULT 43

US-08-629-001A-3
Sequence 3, Application US/08629001A
Patent No. 5858661
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
NUMBER OF SEQUENCES: 139
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 5858661thwestern Hwy.
CITY: Farmington Hills
STATE: Michigan
COUNTRY: US
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: US/08/629,001A
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 2290.00032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 539-5050
TELEFAX: (810) 539-5055
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3056 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-629-001A-3

Query Match 4.3%; Score 81.5; DB 2; Length 3056;
Best Local Similarity 18.5%; Pred. No. 47;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYIKIAL-----KLRQSLFFQNSQSLQRAYSTPYSYVRIIL---QKENKEQ 57
DB 1382 PSHVIKATPAYISNCHTKLKSLILSK-----SPDSYQKILLAIQEAETNN 1431
QY 58 ALARHKCISILEFFKNLLFVHLLSLSKNORECSTDMVSTPFFNRNLWYLLSRFSL 117
DB 1432 VYKGRILKIYHLFVSLLL-----KDKSLGGAWA-----FVLRDVIYTLIH----- 1474
QY 118 WKSVCPRFFLDY-LEAFGLLSDFLDH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166
DB 1475 YINQRPSCIMDVSLRSFSLCCDLLSQVCQTAVTYCKDALENHL--HVIVGTLLIPLVYEQV 1532
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206
DB 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFDPHVFKDLRITQOKIKYSG 1586
QY 207 PHLHPSEFFINMGRLETKVIEKQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266
DB 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRRQLELHKDQVDMRASQDNPQDGINV 1645
QY 267 -----NHEGRK-----AYGAVL 278
DB 1646 KLVNLLQLSKMAINHTEGEVLEAVGSCL 1675

RESULT 44

US-08-874-266-2
Sequence 2, Application US/08874266
Patent No. 5953279
GENERAL INFORMATION:
APPLICANT: Gatti, Richard A.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA MUTATIONS IN THE ATM GENE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874,266
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ways Vensko, Nancy
REGISTRATION NUMBER: 36,298
REFERENCE/DOCKET NUMBER: UCLA006.006A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 24

LENGTH: 3056 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-674-266-2

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 2; Length 3056;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSTPYGYRIIL--QKENKQ 57
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431
QY 58 ALARHKCISILEFFKNLLFVHLISLKNQREGCSTDMAVVSTPFPRNLMYRLLSRPSL 117
DB 1432 VYKGRILKIYHLFVSL--KDKSGLGAWA-----FVLADVIYTLIH----- 1474
QY 118 WKSQCPFFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYYPVSGFVAP----- 166
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206
DB 1533 EVQKQVLDL--KYL-----VIDNKNENIYITIKLDPFPHVFKDLRIQOKIKYRSG 1586
QY 207 PMLHPSEFPINMGRLETKVIEKQALPSKKORIOTLQSNLIAIYRCFWFVSGGLE 266
DB 1587 PSLLEELINH-FLSVSYDALPULTEBGLKDLRQLELHKQVMDIMRASQDNPDGIMV 1645
QY 267 -----NHEGRK---AYGAVL 278
DB 1646 KLVNLLQLSKMAINHTEKEVLEAVGSCL 1675

RESULT 45
US-08-642-274D-3
Sequence 3, Application US/08642274D
Patent No. 6200749
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
FILE REFERENCE: 229000033
CURRENT APPLICATION NUMBER: US/08/642,274D
CURRENT FILING DATE: 1996-05-03
NUMBER OF SEQ ID NOS: 220
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 3
LENGTH: 3056
TYPE: PKT
ORGANISM: Homo sapiens
US-08-642-274D-3

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 4; Length 3056;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSTPYGYRIIL--QKENKQ 57
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431
QY 58 ALARHKCISILEFFKNLLFVHLISLKNQREGCSTDMAVVSTPFPRNLMYRLLSRPSL 117
DB 1432 VYKGRILKIYHLFVSL--KDKSGLGAWA-----FVLADVIYTLIH----- 1474
QY 118 WKSQCPFFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYYPVSGFVAP----- 166
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206

DB 1533 EVQKQVLDL--KYL-----VIDNKNENIYITIKLDPFPHVFKDLRIQOKIKYRSG 1586
QY 207 PMLHPSEFPINMGRLETKVIEKQALPSKKORIOTLQSNLIAIYRCFWFVSGGLE 266
DB 1587 PSLLEELINH-FLSVSYDALPULTEBGLKDLRQLELHKQVMDIMRASQDNPDGIMV 1645
QY 267 -----NHEGRK---AYGAVL 278
DB 1646 KLVNLLQLSKMAINHTEKEVLEAVGSCL 1675

RESULT 46
US-08-952-127-3
Sequence 3, Application US/08952127
Patent No. 6211336
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
APPLICANT: Tagle, Danilo A.
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6211336thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,127
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290,00029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5050
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3056 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-952-127-3

Query Match
Best Local Similarity 4.3%; Score 81.5; DB 4; Length 3056;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYILKIAL-----KLRQSLSLFQNSQSLQRAVSTPYGYRIIL--QKENKQ 57
DB 1382 PSHVIAKATPAVYISNCHKTKLSILEILSK-----SPDSYQKILALICEQALETNN 1431
QY 58 ALARHKCISILEFFKNLLFVHLISLKNQREGCSTDMAVVSTPFPRNLMYRLLSRPSL 117
DB 1432 VYKGRILKIYHLFVSL--KDKSGLGAWA-----FVLADVIYTLIH----- 1474
QY 118 WKSQCPFFLDY-LEAFGLSDPLDH--QAVIKFE--LETHSYYPVSGFVAP----- 166
DB 1475 YINQRPSCIMDVLSRFSCLCDLISQVCQTAVYCKKALENHL--HVIYGTLLPLVYEQV 1532
QY 167 ---HOYLSLLQDRYFPPIASVWRTLDKNFSLTPDLIHDLLGHV----- 206

Db 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVKDLRITQOKIKYSRG 1586
QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266
Db 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDGIMV 1645
QY 267 -----NHEGRK-----AVGAVL 278
Db 1646 KLVNLLQLSKMAINTGEKLEAVGSCL 1675

RESULT 47
US-08-952-014C-3

; Sequence 3, Application US/08952014C
; Patent No. 6265158
; GENERAL INFORMATION:
; APPLICANT: Shiloh, Yosef
; TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
; FILE REFERENCE: GENOMIC ORGANIZATION
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kohn & Associates
; STREET: 30500 No. 6265158thwestern Hwy., Suite 410
; CITY: Farmington Hills
; STATE: Michigan
; COUNTRY: U.S.
; ZIP: 48334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,014C
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohn, Kenneth I.
; REGISTRATION NUMBER: 30,995
; REFERENCE/DOCKET NUMBER: 2290.00028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-539-5050
; TELEFAX: 810-539-5055
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3056 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens

US-08-952-014C-3

Query Match 4.3%; Score 81.5; DB 4; Length 3056;
Best Local Similarity 18.5%; Pred. No. 47; Indels 97; Gaps 17;
Matches 61; Conservative 58; Mismatches 114;

QY 10 PKYLKAL-----KQSLSPFQNSQSLQRAYSTPYSYRIIL---OKENKEQ 57
Db 1382 PSHVIKATFAYISNCHTKLSILELSK-----SPDSYQKILLAIQCAATNN 1431
QY 58 ALARHKICISILEFPKLLFVHLLSLSKNORECSTDMVSTPFNNLWYRLSSRFSL 117
Db 1432 VYKGRILKIYHLFVSLLL-----KDKSLGGGAWA-----FVLRDVIYTLIH----- 1474
QY 118 WKSYPFRFLDY-LEAFGLSLDLH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166
Db 1475 YINQRPSCIMDVLSRSLSCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1532
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206

Db 1533 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVKDLRITQOKIKYSRG 1586
QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266
Db 1587 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDGIMV 1645
QY 267 -----NHEGRK-----AVGAVL 278
Db 1646 KLVNLLQLSKMAINTGEKLEAVGSCL 1675

RESULT 48

US-09-360-416-3

; Sequence 3, Application US/09360416
; Patent No. 6438536
; GENERAL INFORMATION:
; APPLICANT: Richard A. Gatti
; TITLE OF INVENTION: METHODS FOR DETECTION OF ATAXIA
; FILE REFERENCE: TELANGIECTASIA MUTATIONS
; CURRENT APPLICATION NUMBER: US/09/360,416
; CURRENT FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 143
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 3057
; TYPE: PRT
; ORGANISM: Human

US-09-360-416-3

Query Match 4.3%; Score 81.5; DB 4; Length 3057;
Best Local Similarity 18.5%; Pred. No. 47;
Matches 61; Conservative 58; Mismatches 114; Indels 97; Gaps 17;

QY 10 PKYLKAL-----KQSLSPFQNSQSLQRAYSTPYSYRIIL---OKENKEQ 57
Db 1383 PSHVIKATFAYISNCHTKLSILELSK-----SPDSYQKILLAIQCAATNN 1432
QY 58 ALARHKICISILEFPKLLFVHLLSLSKNORECSTDMVSTPFNNLWYRLSSRFSL 117
Db 1433 VYKGRILKIYHLFVSLLL-----KDKSLGGGAWA-----FVLRDVIYTLIH----- 1475
QY 118 WKSYPFRFLDY-LEAFGLSLDLH--QAVIKFFE--LETHFSYYPVSGFVAP----- 166
Db 1476 YINQRPSCIMDVLSRSLSCDLLSQVCQTAVTYCKDALENHL--HVIVGTLPVYEQV 1533
QY 167 ---HQYLSLLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLGHV----- 206
Db 1534 EVQKQVLDLL--KYL-----VIDNKNENLYITIKLLDPFPHVVKDLRITQOKIKYSRG 1587
QY 207 PWLLHPSFSEFFINMGRLEFTKVIKQVQALPSKQRIOTLQSNLIAIVRCFWFTVESGLIE 266
Db 1588 PFSLEBINH-FLSVSVYDALPLTRLEGLKDLRQLEHLKQDMQVIMRASQDNPDGIMV 1646
QY 267 -----NHEGRK-----AVGAVL 278
Db 1647 KLVNLLQLSKMAINTGEKLEAVGSCL 1676

RESULT 49

US-08-540-804-18

; Sequence 18, Application US/08540804
; Patent No. 5919666
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

Fri Jan 10 10:56:11 2003

us-09-438-185a-1047.ra1

Page 26

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,804
FILING DATE: 11-OCT-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/521,872
FILING DATE: 21-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,265
FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-540-804-18

Query Match

Best Local Similarity 23.0%; Score 81; DB 2; Length 323;
Matches 56; Conservative 31; Mismatches 71; Indels 86; Gaps 12;

QY 122 CRRFLDYI-EAFGLSDFL--DHOAV--IKFPELETHFSYVSGFVAPHQYLSLQDR 176
DB 134 CQYIRTLVSEARTLWPEFIPDPKYTEFEFYLLLEESY-----LIYHNPYQSLKQ-- 186
QY 177 YFPIASVNR-----TLDKNPSLTPDLIHDLGHVWMLHPSSEFFINMGRLETKVIE 230
DB 187 ---IVQVLKQPPQITLSSDQLQNCWSLINDSYINDVHLLYPRH----- 227
QY 231 KVOALPSKKQRIQTLSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFI 290
DB 228 -----IIVACLFTI-----SIHGKPTKSSLASAASE----- 256
QY 291 DNRVVLPLEDQIIRLPNTSTPOETLFSIRHPELVELTSKLEMMLDQGLIESIPLYNQ 350
DB 257 -----AIRDPKSSSPVOIAFNRMASLVD---LBEVMDT-TQEQITLYDH 299
QY 351 -EKY 353
DB 300 WDKY 303

RESULT 50
US-08-218-265-18
Sequence 18, Application US/08218265
Patent No. 5922585
GENERAL INFORMATION:
APPLICANT: Young, Richard A.
APPLICANT: Thompson, Anthony J.
APPLICANT: Koleske, Craig M.
TITLE OF INVENTION: No. 5822585el Factors Which Modify Gene
NUMBER OF SEQUENCES: 35
TRANSCRIPTION AND METHODS OF USE THEREOF
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/218,265
FILING DATE: 25-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI94-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-218-265-18

Query Match

Best Local Similarity 23.0%; Score 81; DB 2; Length 323;
Matches 56; Conservative 31; Mismatches 71; Indels 86; Gaps 12;

QY 122 CRRFLDYI-EAFGLSDFL--DHOAV--IKFPELETHFSYVSGFVAPHQYLSLQDR 176
DB 134 CQYIRTLVSEARTLWPEFIPDPKYTEFEFYLLLEESY-----LIYHNPYQSLKQ-- 186
QY 177 YFPIASVNR-----TLDKNPSLTPDLIHDLGHVWMLHPSSEFFINMGRLETKVIE 230
DB 187 ---IVQVLKQPPQITLSSDQLQNCWSLINDSYINDVHLLYPRH----- 227
QY 231 KVOALPSKKQRIQTLSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFI 290
DB 228 -----IIVACLFTI-----SIHGKPTKSSLASAASE----- 256
QY 291 DNRVVLPLEDQIIRLPNTSTPOETLFSIRHPELVELTSKLEMMLDQGLIESIPLYNQ 350
DB 257 -----AIRDPKSSSPVOIAFNRMASLVD---LBEVMDT-TQEQITLYDH 299
QY 351 -EKY 353
DB 300 WDKY 303

Search completed: January 9, 2003, 17:00:47
Job time : 42 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 14:44:28 ; Search time 81 Seconds
(without alignments)
920.953 Million cell updates/sec

Title: US-09-438-185a-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKIALKL.....ESIPLYNQEKYLSGFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	237.5	12.6	491	13	Q9PU40
2	224.5	11.9	497	5	Q96370
3	219	11.6	579	5	Q24000
4	219	11.6	579	5	Q8SY95
5	215.5	11.4	453	11	Q91WV1
6	213.5	11.3	452	4	Q8TEY0
7	213.5	11.3	452	4	Q8TC14
8	206.5	10.9	244	2	Q9AG78
9	206.5	10.9	522	5	Q23438
10	206.5	10.9	532	5	Q9XZD1
11	181.5	9.6	457	5	Q9XYQ5
12	178.5	9.4	450	5	Q96947
13	178.5	9.4	555	5	Q9W0K2
14	177	9.4	495	2	Q9PDC3
15	157	8.3	438	5	Q17498
16	135.5	7.2	250	6	Q95LQ6

17	109.5	5.8	1253	10	Q9SCZ3	Q9scz3 arabidopsis
18	106.5	5.6	506	8	Q9GF63	Q9gf63 arabis alpi
19	106	5.6	129	13	Q42428	Q42428 lates calca
20	106	5.6	504	8	Q9GF55	Q9gf55 cardamine a
21	106	5.6	504	8	Q9GF31	Q9gf31 cardamine p
22	106	5.6	504	8	Q9GF30	Q9gf30 cardamine r
23	104	5.5	513	8	Q95F33	Q95f33 muhlenbergi
24	104	5.5	513	8	Q95F29	Q95f29 enneapogon
25	103.5	5.5	513	8	Q95F39	Q95f39 monanthochl
26	103.5	5.5	1782	5	Q9VP19	Q9vp19 drosophila
27	103	5.5	504	8	Q9GF41	Q9gf41 aubrieta de
28	103	5.5	513	8	Q9TI88	Q9tib8 bouteloua c
29	103	5.5	513	8	Q95F60	Q95f60 buchloe dac
30	103	5.5	513	8	Q95F50	Q95f50 schedonnard
31	103	5.5	513	8	Q95F38	Q95f38 reederochlo
32	102.5	5.4	1154	5	Q9VM23	Q9vm23 drosophila
33	102	5.4	506	8	Q9GF64	Q9gf64 arabis alpi
34	101.5	5.4	519	8	Q9MV23	Q9mv23 oryza meyer
35	101.5	5.4	1615	5	Q8T2D4	Q8t2d4 dictyosteli
36	101	5.3	635	17	Q972M5	Q972m5 sulfolobus
37	101	5.3	2042	5	Q9W1C5	Q9w1c5 drosophila
38	100.5	5.3	102	11	P97517	P97517 phodopus su
39	100.5	5.3	2212	12	Q9DD11	Q9dgd1 ebola virus
40	99.5	5.3	513	8	Q9TI97	Q9ti97 sporobolus
41	99	5.2	513	8	Q95F61	Q95f61 asagopogon c
42	99	5.2	513	8	Q95F52	Q95f52 symnagogon
43	99	5.2	924	11	Q54321	Q54321 rattus norv
44	99	5.2	1078	17	Q8TFH6	Q8tfh6 methanosarc
45	98.5	5.2	513	8	Q9TIC0	Q9tico zeuglites pi
46	98.5	5.2	584	16	Q9PLZ8	Q9plz8 campylobact
47	98.5	5.2	945	10	Q94DZ7	Q94dz7 oryza sativ
48	98	5.2	924	11	Q9D4H1	Q9d4h1 mus musculu
49	98	5.2	1887	5	Q9GZG6	Q9gzg6 caenorhabdi
50	97.5	5.2	239	5	Q9NJQ3	Q9njq3 lymaea sta
51	97.5	5.2	504	8	Q9GF36	Q9gf36 lepidium ca
52	97.5	5.2	519	8	Q9MV31	Q9mv31 oryza alta
53	97.5	5.2	519	8	Q9MV30	Q9mv30 oryza latif
54	97.5	5.2	519	8	Q9MV29	Q9mv29 oryza grand
55	97.5	5.2	946	10	Q94E00	Q94e00 oryza sativ
56	97	5.1	347	4	Q13786	Q13786 homo sapien
57	97	5.1	504	8	Q9GF53	Q9gf53 arabis proc
58	97	5.1	504	8	Q9GF28	Q9gf28 crucihimala
59	97	5.1	513	8	Q95F32	Q95f32 muhlenbergi
60	97	5.1	519	8	Q9MV25	Q9mv25 oryza ridle
61	97	5.1	519	8	Q9MV24	Q9mv24 oryza longi
62	96.5	5.1	363	8	Q9GHV8	Q9ghv8 caryodaphno
63	96	5.1	515	8	Q9GHD7	Q9ghd7 helionopsis
64	96	5.1	2059	12	Q56278	Q56278 human herpe
65	95.5	5.1	355	8	Q9GHT3	Q9ght3 mezilaurus
66	95.5	5.1	355	8	Q9GHM9	Q9ghm9 sassafras t
67	95.5	5.1	536	16	P96714	P96714 bacillus su
68	95.5	5.1	993	8	Q94QR6	Q94qr6 pleurotus o
69	95	5.1	2185	12	Q9E948	Q9e948 apple stem
70	95	5.0	502	8	Q9GF40	Q9gf40 arabis blep
71	95	5.0	504	8	Q9GF38	Q9gf38 crucihimala
72	95	5.0	970	10	Q22941	Q22941 arabidopsis
73	95	5.0	1147	5	Q9NSY5	Q9nsy5 caenorhabdi
74	94.5	5.0	245	12	Q91MT1	Q91mt1 lumpy skin
75	94.5	5.0	355	8	Q9GHM2	Q9ghm2 tambourissa
76	94.5	5.0	513	8	Q95F55	Q95f55 chloris tru
77	94.5	5.0	519	8	Q9MV37	Q9mv37 oryza offic
78	94.5	5.0	519	8	Q9MV36	Q9mv36 oryza eichi
79	94.5	5.0	519	8	Q9MV35	Q9mv35 oryza rhizo
80	94.5	5.0	699	16	Q9PND4	Q9pnd4 campylobact
81	94.5	5.0	212	12	Q91DD4	Q91dd4 ebola virus
82	94	5.0	293	17	Q9ZMC6	Q9zmc6 helicobacte
83	94	5.0	317	17	Q5067	Q5067 pyrococcus
84	94	5.0	504	8	Q9GF62	Q9gf62 barbarea vu
85	94	5.0	504	8	Q9GF51	Q9gf51 arabidopsis
86	94	5.0	504	8	Q9GF33	Q9gf33 microthlas
87	94	5.0	513	8	Q9TIA8	Q9tia8 perotlis rar
88	94	5.0	513	8	Q95F36	Q95f36 tuctoria gr
89	94	5.0	618	16	Q9X0Q2	Q9x0q2 thermotoga

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN (1)
 RC SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stagleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Fisse E.,
 RA George R., Gonzalez M., Guatin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celisner S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY071698; AAL49320.1; -
 SO SEQUENCE 579 AA; 66012 MW; 7700F94E436479E2 CRC64;

Query Match 11.6%; Score 219; DB 5; Length 579;
 Best Local Similarity 28.6%; Pred. No. 28-10; Mismatches 94; Indels 54; Gaps 8;
 Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

QY 128 DYLEAFGLSD---FLDH---QAVIKFELETHRSYVSGFAPRHOYSLQDXYFP 179
 DB 328 EYRAHFOKLODEQIFVERRLPQLOEMDFLRKRTGSLRPAAGLTADELASLAFRIG 388
 QY 180 IASVWRITDKDNFSLTPDLIHDLGVPMILHPSSEPTINMGRLFTKYIEKVALPSKK 239
 DB 389 STQYRVHNSPHTPEPSHGLHMPILADPSFPOFQOIG-----LASLGASD 439
 QY 240 QRIQTLQSNLIAVRCFMPFVSGLIENHGRKAYCAVLISSPOLGHAFLD----- 291
 DB 440 EEIEKLST-----VYMFVEFGLCCKEHOIAYAGALLSYGELHLAIDCKEHRPE 492
 QY 292 -NVRVLE-----LDQIRLPNTSTPQETLFSIRHPD-ELYELT 330
 DB 493 PASTVAVQYQOQEXQPIYYAESFEDAKDKFRRWSTMRPFVEFRN-PTHERVEVLDSV 551
 QY 331 SKLEWMLDQGLLESIPLYN 349
 DB 552 DKLETLVHQMTETILHLTN 570

RESULT 5
 O91WV1 PRELIMINARY; PRT; 453 AA.
 AC O91WV1.
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Phenylalanine hydroxylase.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OK NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RC Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC013458; AAH13458.1; -
 DR InterPro: IPR001273; Aaa_hydroxylase.
 DR InterPro: IPR002912; ACT.
 DR Pfam: PF01842; ACT; 1.
 DR Pfam: PF00351; bioprotein H; 1.
 DR PRINTS: PRO0372; FWHYDRXLASE.
 DR TIGRFAMs: TIGR01268; Pheahydroxylase.
 DR PROSITE: PS00367; BIOPROTEIN HYDROXYL; UNKNOWN 1.
 SO SEQUENCE 453 AA; 51899 MW; 551P181FA59DA5B CRC64;

Query Match 11.4%; Score 215.5; DB 11; Length 453;
 Best Local Similarity 27.8%; Pred. No. 2.9e-10;
 Matches 64; Conservative 40; Mismatches 105; Indels 21; Gaps 5;

QY 104 RNLM---YLLSRFSLMKSYPREFLDYLAFLGLSDFLDH-QAVIKFELETHRSY 158
 DB 184 KRTWGTVFRTKLKALYKTHACVHNHNPFLLEKFCGFREDNLPQLEDSQGLQCTGFRUR 243
 QY 159 PVSGFVAPHOYSLQDRYFPPIASVWRITDKDNFSLTPDLIHDLGVPMILHPSSEFF 218
 DB 244 PVAGLSSRFGLGLAFRVFHCOTYIRHSGSKMPTPEPDICHELHGVPLPSDRSFAQS 303
 QY 219 INMGRLFTKYIEKVALPSKKORIQTQSNLIAVRCFMPFVSGLIENHGRKAYCAVLI 278
 DB 304 QEID-----LASLGAPDEYIEKLAT-----YMFVEFGLCCKEDSDIKAYAGLI 347
 QY 279 ISSPOLGHAFLDNVAVLPLELDQIRLPNTSTPQETLFSIRFDELVE 328
 DB 348 LSSFGLQYCLSDKPKLPLELEKTAQCEYVTEFQPLYYVAESFNDAKE 397

RESULT 6
 O8TEY0 PRELIMINARY; PRT; 452 AA.
 AC O8TEY0.
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Phenylalanine hydroxylase.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Konecki D.S., Licher-Konecki U.;
 RT "Completion of the sequence of PAH, a model disease gene."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF404777; AAL78816.1; -
 SO SEQUENCE 452 AA; 51861 MW; DEF9DB9D6B8C800C CRC64;

Query Match 11.3%; Score 213.5; DB 4; Length 452;
 Best Local Similarity 26.6%; Pred. No. 4.4e-10;
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLM---YLLSRFSLMKSYPREFLDYLAFLGLSDFLDH-QAVIKFELETHR 155
 DB 184 KRTWGTVFRTKLKALYKTHACVHNHNPFLLEKFCGFREDNLPQLEDSQGLQCTGFR 240
 QY 156 SYVSGFVAPHOYSLQDRYFPPIASVWRITDKDNFSLTPDLIHDLGVPMILHPSFS 215
 DB 241 RLRVAGLSSRFGLGLAFRVFHCOTYIRHSGSKMPTPEPDICHELHGVPLPSDRSFA 300
 QY 216 EPTINMGRLFTKYIEKVALPSKKORIQTQSNLIAVRCFMPFVSGLIENHGRKAYG 275
 DB 301 QFSQEID-----LASLGAPDEYIEKLAT-----YMFVEFGLCCKQSDSIKAYG 344
 QY 276 AVLISSPOLGHAFLDNVAVLPLELDQIRLPNTSTPQETLFSIRFDELVE 328
 DB 345 AGLISSFGLQYCLSEKPKLPLELEKTAQCEYVTEFQPLYYVAESFNDAKE 397

RESULT 7
 O8TC14 PRELIMINARY; PRT; 452 AA.
 AC O8TC14.
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Phenylalanine hydroxylase.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OK NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.

```

RC TISSUE=LIVER;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026251.1; RAH26251.1;
SQ SEQUENCE 452 AA; 51790 MW; 018B9629BBD640D CRC64;

Query Match 11.3%; Score 213.5; DB 4; Length 452;
Best Local Similarity 26.6%; Pred. No. 4.4e-10;
Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLMYRLSSRFSLSKSCYCRFFLDYLAFLGSLSDFLDH-----QAVIKFELETHF 155
DB 184 KKTGTVFKTLKSLYKTHA---CYEYNHFFLEKYCGFHEDNIPQLEDVQSPLQCTGTF 240
QY 156 SYYPVSGFVAPHQYLSLLQDRYFPFIASVWRTLKDNFSLTPDLIHLGLHGVWPLHPSPS 215
DB 241 RLRPVAGLLSRDFLGLAFRVFCHCTQVIRHGSKPMYTPEDICHELLGHVPLFSDRSPA 300
QY 216 EFFINMGRLEFTKVIKQVLPKSKQRIQTLOSNIATVRCFWFTVESGLIENHEGRKAYG 275
DB 301 QFSQIEG-----LASLGAPDEYIEKLTAT-----IYWFTVEFLCKQGDSTKAYG 344
QY 276 AVLSSSQEELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDELVE 328
DB 345 AGLSSFCGLQYCLSEKPKLLPLEKTAIQNYTVTEPQPLYYVAESFNDAKE 397

RESULT 8
QY 09AG78 PRELIMINARY; PRT; 244 AA.
DI 01-JUN-2001 (TEMBLrel. 17, Created)
DI 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DI 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Amino acid hydroxylase.
OS Streptomyces verticillus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=29309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC15003;
RA Du L., Sanchez C., Chen M., Edwards D.J., Shen B.;
RT "A locus encoding nonribosomal peptide synthetase and polyketide
RT synthase functions in the bleomycin producer Streptomyces verticillus
RT ATCC15003."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF340166; AAK27471.1;
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioprotein H; 1.
SQ SEQUENCE 244 AA; 27012 MW; 31962679A28916C7 CRC64;

Query Match 10.9%; Score 206.5; DB 2; Length 244;
Best Local Similarity 23.9%; Pred. No. 8.2e-10;
Matches 59; Conservative 48; Mismatches 123; Indels 17; Gaps 4;

QY 97 VSTPP--FNNMLYRLSSRFSLSKSCYCRFFLDYLAFLGSLSDFL-DHQAVIKFELET 153
DB 9 VQEPFGVQDVQVWRLHYRQAEVLVPAAPAPPLYLEGALLGLTDPDGLQGVNRLSAVS 68
QY 154 HFSYYPVSGFVAPHQYLSLLQDRYFPFIASVWRTLKDNFSLTPDLIHLGLHGVWPLHPSPS 213
DB 69 GWTCAVDGLVKQPEFDMLLHRRFPVTSRMPELFAFLPDLFDLFGHGPLYLAHAR 128
QY 214 FSEFFINMGRLEFTKVIKQVLPKSKQRIQTLOSNIATVRCFWFTVESGLIENHEGRKA 273
DB 129 TADLYQOEFGRIQV-----CAGRPEDPALLQGLI-----WSTLETGLIRTPGLRA 174
QY 274 YCAVLISSQEELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDELVEITSL 333
DB 175 LGGAILSSADEIRQCLDPAFCVFPDPVVRVATVDILRLQSRVFAVEDELEETESALDL 234
QY 334 EWLMDQG 340

```

```

DB 235 DALVTG 241
RESULT 9
QY 023438 PRELIMINARY; PRT; 522 AA.
DI 01-NOV-1996 (TEMBLrel. 01, Created)
DI 01-OCT-2001 (TEMBLrel. 18, Last sequence update)
DI 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical 59.8 kDa protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2016(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Taich A.;
RT "The sequence of C. elegans cosmid ZK1290."
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U21308; AAB93319.2;
DR HSSP; P04176; 1PHZ.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioprotein H; 1.
DR PRINTS; PR00372; FYMHYDRLASE.
DR TIGRFAMs; TIGR01270; Tfp_5_monoox; 1.
KW Hypothetical protein.
KW SEQUENCE 522 AA; 59781 MW; B6205C4B932C21FA CRC64;

Query Match 10.9%; Score 206.5; DB 5; Length 522;
Best Local Similarity 25.2%; Pred. No. 2.1e-09;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNLM-----YRLSSRFSLSKSCYCRFFLDYLAFLGSLSDFLDH-----QAVIKFELETHF 155
DB 241 KKTGTVFKTLKSLYKTHA---CYEYNHFFLEKYCGFHEDNIPQLEDVQSPLQCTGTF 297
QY 156 SYYPVSGFVAPHQYLSLLQDRYFPFIASVWRTLKDNFSLTPDLIHLGLHGVWPLHPSPS 215
DB 298 RVRPVAGLLSRDFLGLAFRVFCHCTQVIRHGSKPMYTPEDICHELLGHVPLFSDRSPA 357
QY 216 EFFINMGRLEFTKVIKQVLPKSKQRIQTLOSNIATVRCFWFTVESGLI-----265
DB 358 QFSQIEG-----LASLGASEEDLKLATL-----YFFSIEFGLSSDDAADSPVK 401
QY 266 ---ENHEGRKAYGAVLISQPELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRH 322
DB 402 ENGNSHERFVKYVAGLLSAGSLQHAVESGATIRFPDPRVQECILITTFQSAYFYTRN 461
QY 323 FDELVELTSKLEWMLDQGLLESLIPLYNQEKYLSGFEVL 360
DB 462 FEEAQQ---KLRFMTNNMKRPFIYRN--PYTESVEVL 494

RESULT 10
QY 09XZ01

```

```

ID Q9XZD1 PRELIMINARY; PRT; 532 AA.
AC Q9XZD1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE tryptophan hydroxylase.
GN TPH-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sze J.Y., Ruvkun G.;
RT "tpb-1 encodes a C. elegans tryptophan hydroxylase.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DDJ databases.
DR EMBL; AF135186; AAD30115.1; -.
DR HSSP; P04176; 1PH2
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01270; tpd_5_monoox; 1.
SQ SEQUENCE 532 AA; 60863 MW; 8AC6A51C7DD0121F CRC64;

Query Match 10.9%; Score 206.5; DB 5; Length 532;
Best Local Similarity 25.2%; Pred. No. 2.1e-09;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNLW---YRLSSRFLSKSYCPFFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155
DB 251 RTWGIITRYRKLK---ELKKKACCKOFLDNFELLEHRGCGSENNIPOLBIDICKELAKTGF 307
QY 156 SYYPVSGFVAPHQYLSLQDRYFPFIASVMRTLDKNFSLTPDLIDHLDGHVWMLHPSFS 215
DB 308 RVRPAGYLSADFLAGLAVYVFCTQYVRHNDPPTPEBPVHELMGHMALFADPDFA 367
QY 216 EFFIMNGRLFTKVEIKVOALPSKKORIQTQSNLIAIVGCFMFTVSGLI----- 265
DB 366 QFSQELG-----LASIGASEEDDKLATL-----YFSTIEFGLSDGDADSPVK 411
QY 266 ---ENHEGRKAYGAVILSSPOELGAFIDNVRVLPLEDOIIRLPNTSTQETLFSIRH 322
DB 412 ENGSHNERFKYAGAGLSSAGELGHAIVESATITIRPDPRVVEQCLITTFQSAVYTRN 471
QY 323 FDELVELTSKLEWMLDGLLESIPLYNQEKYLSGFVYL 360
DB 472 FEEAQQ---KLRFMTNNMKRPFIIVRN--PYTESVEVL 504

RESULT 11
ID Q9XYO5 PRELIMINARY; PRT; 457 AA.
AC Q9XYO5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Phenylalanine hydroxylase (EC 1.14.16.1).
GN K08F8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Loei C.M., Davidson B., McKerrow J.;
RT "A phenylalanine hydroxylase gene from the nematode Caenorhabditis
RT elegans is expressed in the hypodermis.";
RL J. Neurogenet. 0:0-0(1999).
DR EMBL; AF119388; AAD31643.1; -.
DR HSSP; P00439; 4PAH.
DR TIGRfams; TIGR01273; Phe4hydrox. tet; 1.
DR InterPro; IPR001273; Aaa_hydroxylase.
SQ SEQUENCE 457 AA; 52189 MW; 7A573B84B9BE6FC CRC64;

```

```

DR InterPro; IPR000179; Cyt_b_6.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01268; Phe4hydrox. tet; 1.
DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; UNKNOWN_1.
KM Oxidoreductase.
SQ SEQUENCE 457 AA; 52189 MW; 7A573B84B9BE6FC CRC64;

Query Match 9.6%; Score 181.5; DB 5; Length 457;
Best Local Similarity 21.8%; Pred. No. 2.4e-07;
Matches 83; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 14 LKIALKTRQSLSPFONSQSLORAY----STPYSTYRIILQKENEKQALARKICISILE 69
DB 81 LKTVRLKELSLST---SNKKLRRLFKTGIPKT-----KQKDSVWPFQKINDIDQ 130
QY 70 FPKNLL-----FVHLISLKNQEGCSTDMA-----VSTPFENRL--WYRL 110
DB 131 FANRLISYGAEILDADHPGFKMDTYRERKFPADIAFNFKQKDKIPITYDEBIATWTV 190
QY 111 LSSRFSLW-KSYCPFFLDYLEAFGLSD-----FLDHQAVIKFPELETHFSYYPVS 161
DB 191 YNELVTWVPKMACQEF--NYI--FELLQONCGPDRIPOLQVSDFLKDCCTGYTRPA 246
QY 162 GFVAPHQYLSLQDRYFPFIASVMRTLDKNFSLTPDLIDHLDGHVWMLHPSSEPFIM 221
DB 247 GLSPRDFLAGMAFRVFNHSTGYIRHNSAPKYPPEPDICHELGLHVPFADVEPAQPSQRI 306
QY 222 GRLE---TVIEKVOALPSKKORIQTQSNLIAIVGCFMFTVSGLIENHEGRKAYGAV 277
DB 307 GLASLAPDQVIEIKLATL-----VFTIEFGLCQDDDEKRAYGAG 346
QY 278 LSSPOELGAFIDNVRVLPLE--LDQIRLPNTSTQETLFSIRH 317
DB 347 LLSFGLQYALSDKREVDPRPVCCTKPTITREYQKFLFLAESPAKAKLKSWAAT 406
QY 318 ---FSIRH--FDELVELTSK 333
DB 407 NRPFQIRKYNAYTORVAILDKV 427

RESULT 12
ID Q96947 PRELIMINARY; PRT; 450 AA.
AC Q96947;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Phenylalanine hydroxylase.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodiidae; Geodia.
NCBI_TaxID=6047;
RN [1]
RP SEQUENCE FROM N.A.
RA Wiene M., Kozioł C., Barclay R., Wheller W.;
RT "Phenylalanine hydroxylase from the sponge Geodia cydonium:
RT Implication for allorecognition and evolution of aromatic amino acids
RT hydroxylases.";
RL Dav. Comp. Immunol. 22:469-478(1998).
DR EMBL; Y16353; CAA76184.1; -.
DR HSSP; P00439; 2PAH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR InterPro; IPR002912; ACT.
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioperlin_H.1.
DR PRINTS; PR00372; FYMHYDRXLASE.
DR TIGRfams; TIGR01268; Phe4hydrox. tet; 1.
SQ SEQUENCE 450 AA; 51204 MW; 691880218BB9725D CRC64;

```


Query Match 9.4%; Score 178.5; DB 5; Length 450;
Best Local Similarity 25.8%; Pred. No. 4.3e-07;
Matches 55; Conservative 43; Mismatches 80; Indels 35; Gaps 6;
QY 107 WYRLSSRFLSWKSYCPFRFDLYLEAFGLSDFLDH-----QAVIKFELETHFSY 158
DB 186 WRTITNLVLFPTHACK--EHNHVPFLQENCGYREDNIPOLEVSQYLSCTGFLRL 242
QY 159 PVSGFVAPHVLSLQDRYFPFIASVMTLDKDNFSLTDLHDHGLGHVPLLHPSSEFP 218
DB 243 PVAGLLSRDFLAGLAFRVPHSVQYIRHVSNQYPTPDVCHLGHVRCV-ILFAQFS 301
QY 219 INMGRLTKVIEKQVLPKQKRIQTLQSNLIAVRCFWFVSGLIENHGRKAYGAVL 278
DB 302 QEIG-----LASGAPEEVVOGLAT-----LWFTIEFLCKQDGQTKYAGAGL 345
QY 279 ISSQELGHAFINVRVLPLELDQIIRLPFNTS 311
DB 346 ISSGELQYCLSRKPEVRPLD-----PFKTS 371

RESULT 13
QY 09W0K2 PRELIMINARY; PRT; 555 AA.
AC Q9W0K2;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE CG9122 protein.
OS CG9122.
GN Drosophila melanogaster (Fruit fly).
OC Sukariyota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Ephygryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephygryota; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt R.G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelist C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lamel P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Malshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkiam I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL; AE003470; AAF47444.1; -
DR HSSP; P04176; IPHZ.
DR PLYBase; FB00035187; CG9122.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR InterPro; IPR02912; ACT.
DR Pfam; PF01842; ACT.1.
DR Pfam; PF00351; bioprotein H; 1.
DR PRINTS; PR00372; FWHYDLALASE.
DR TRNFS; TIGR01270; TIF_5_monomer; 1.
DR PROSITE; PS00367; BIOPROTEIN HYDROXYL; 1.
SQ SEQUENCE 555 AA; 61530 MW; 60D527FEB1FA791C CRC64;

Query Match 9.4%; Score 178.5; DB 5; Length 555;
Best Local Similarity 32.0%; Pred. No. 5.6e-07;
Matches 47; Conservative 28; Mismatches 55; Indels 17; Gaps 3;

QY 143 QAVIKFELETHFSYYPVSGFVAPHVLSLQDRYFPFIASVMTLDKDNFSLTDLHDH 202
DB 284 QDVSVYLKRTKQFLRPVAGYLSRDLGLAFRVHCTQYIRHSSDPFYTPEDCCHEL 343
QY 203 LGHVPLLHPSSEFPFINMRLFTKVIKQVLPKQKRIQTLQSNLIAVRCFWFVSG 262
DB 344 LGHNPLLANSFSAQFSQEI-----LASGASDADIEKLATL-----YFTVVF 387
QY 263 GLIENHGR-KAYGAVLISPOELGHA 288
DB 388 GLCKQADSTFKVYGAGLLSSVLAELQHA 414

RESULT 14
QY 09FDC3 PRELIMINARY; PRT; 495 AA.
AC Q9FDC3;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE PAH.
GN PAH.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21359321; PubMed=11466282;
RA Sun H., Shi W.;
RT "Genetic Studies of mrp, a Locus Essential for Cellular Aggregation
and Sporulation of Myxococcus xanthus.";
RL J. Bacteriol. 183:4786-4795(2001).
DR EMBL; AF285263; AAR99327.1; -
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioprotein H; 1.
SQ SEQUENCE 495 AA; 53213 MW; CD82B52F8A613AEB CRC64;

Query Match 9.4%; Score 177; DB 2; Length 495;
Best Local Similarity 23.8%; Pred. No. 6.5e-07;
Matches 69; Conservative 47; Mismatches 128; Indels 46; Gaps 6;

QY 91 STDMAVSTPFFNRNLWYRLSSRFLSWKSYCPFRFDLYLEAFGLSDFLDHQAVIKFFE 150
DB 31 SRDQAV-----WRNIIGLRGLHADKAHPVLEGLGATGIGSECTIPSLDENNEKL 80
QY 151 LETHFSYYPVSGFVAPHVLSLQDRYFPFIASVMTLDKDNFSLTDLHDHGLGHVPLL 210
DB 81 ARLCMACVGVGRFIPAVFTELQAGVLAIAADIRTHEIETYPADIVHESAGAPIIA 140
QY 211 HPFSSEFPINWGLFTKVIKQV-----QAL-----PSKORIOTLQSNLIA--- 251
DB 141 NRRVAEYLKAGLVGFKATSVEDQAVFATNLSVVKEDPDASEEAAHAQARLEASA 200

QY 252 -----IVRCFWFTVESGLIENHGRKAYGLISSPOLGAFIDNVRVPLELD 301
 DB 201 SRRVSESTRASRLVMTAEVGLGVASPRIVYAGLFTSTIEBACHLTPVAKLLLSV- 259
 QY 302 QIIRLPNTSTPQETLSIRHPELVELTSKLEWML-----DQGLSESI 345
 DB 260 ACADMDYDITRMQPOLFVARDPEHLFEVLAFEESTLSMKRGDPGLTEAL 309

RESULT 15

ID 017498 PRELIMINARY; PRT; 438 AA.
 AC 017498;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE Phenylalanine hydroxylase (EC 1.14.16.1) (Fragment).
 GN PAH.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 CC Branchiostoma.
 CX NCBI_TaxID=7739;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paton S.J., Luke G.N., Holland P.W.H.;
 RT "Complex history of a chromosomal paralog region: insights from
 RT genes.";
 RL Mol. Biol. Evol. 15:1373-1380 (1998).
 DR EMBL; AJ001677; CAA04917.1; -.
 DR HSSP; P04176; 1PH2.
 DR InterPro; IPR001273; Aaa_hydroxylase.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR002119; Histone_H2A.
 DR Pfam; PF01842; ACT; 1.
 DR Pfam; PF00351; bioprotein_H; 1.
 DR TIGRFAMs; TIGR01268; Phehydrox_tetr; 1.
 DR PROSITE; PS00046; HISTONE_H2A; UNKNOWN_1.
 KW Oxidoreductase.
 FT NON_TER
 FT SEQUENCE 438 AA; 4958 MW; 1619297DBDBF5EE7 CRC64;

Query Match 8.3%; Score 157; DB 5; Length 438;
 Best Local Similarity 24.1%; Pred. No. 2.9e-05;
 Matches 64; Conservative 37; Mismatches 123; Indels 36; Gaps 8;

QY 105 NLMYRLSSRSFSLMKSYCPR-----FIDYIEAFGLSDPLDH-QAVIKFELETHFSYYP 159
 DB 170 NIMROITFELARKLPTDACKENHNVFPLMENGCPREDNIPQLEDVSNLKDCTGFILRP 229
 QY 160 VSGFVAHQVYSLIQDRYFPPLASVMRTLDKONSLTPDLIHLLGHVWMLHPSPSEFFI 219
 DB 230 VAGLLSSRDPLAGLAFPVFHSSTOYIRHSHKPLTPEPDVCHLLGHAPLPADPSFAQPSQ 289
 QY 220 NMGRLTKVIEKQVALPSKQRIQTOGSLNIAIVRCFWFTVESGLIENHGRKAYGAVLI 279
 DB 290 ELG-----LASTGAPDDFWKATL-----YMPTEVFGICRODGEKAYGRDCC 333
 QY 280 SSPQELGAFIDNVRVPLELDQIIRLPNTSTPQETLSIRHPELVELTSKLEWMLDQ 339
 DB 334 HRSSECRCLTDKPDIDRPFEBEKISVTKYPIITEYQIYVAVDSFODAKEKVR--TWS--- 388
 QY 340 GLLESIFL-----YNOEKYLSGFEVL 360
 DB 389 ---HSIPRPSVHN--PYQSVAIL 409

RESULT 16

ID 095106 PRELIMINARY; PRT; 250 AA.
 AC 095106;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE Tyrosine hydroxylase (EC 1.14.16.2) (Fragment).
 GN TH.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 CX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=THOROUGHRED; TISSUE=ADRENAL MEDULLA;
 RA Sato F., Matsuta M., Hasegawa T.;
 RT "Molecular cloning of cDNA for equine tyrosine hydroxylase."
 RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB071421; BAB68123.1; -.
 DR InterPro; IPR001273; Aaa_hydroxylase.
 DR Pfam; PF00351; bioprotein_H; 1.
 DR PROSITE; PS00367; BIOPROTEIN_HYDROXYL; UNKNOWN_1.
 KW Oxidoreductase.
 FT NON_TER
 FT NON_TER
 FT SEQUENCE 250 AA; 28548 MW; 293BCED98F39A6B CRC64;

Query Match 7.2%; Score 135.5; DB 6; Length 250;
 Best Local Similarity 28.2%; Pred. No. 0.001;
 Matches 35; Conservative 25; Mismatches 53; Indels 11; Gaps 2;

QY 107 MYRLSSRSFSLMKSYCPRFIDYIEAFGLSDPLDH-----QAVIKFELETHFSYYP 158
 DB 124 WKEVYTTKLGLVYTHACR--EHLFAFELRLRFGYREDNIPQLEDVSRFLKERSGFQUR 180
 QY 159 VSGFVAHQVYSLIQDRYFPPLASVMRTLDKONSLTPDLIHLLGHVWMLHPSPSEFFI 218
 DB 181 PVAGLLSSRDPLAGLAFPVFHSSTOYIRHSHKPLTPEPDVCHLLGHAPLPADPSFAQPSQ 240
 QY 219 INMG 222
 DB 241 QDIG 244

RESULT 17

ID 098C23 PRELIMINARY; PRT; 1253 AA.

AC 098C23;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE Disease resistance-like protein.
 GN F2E013.200.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
 CX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Delaney M., Berger C., Cooke R., Grellet F., Landie M., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Arabidopsis sequencing project;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL133452; CAB63020.1; -.
 DR InterPro; IPR000767; Disease_resist.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR002182; NB-ARC.
 DR InterPro; IPR00157; TIR_domain.
 DR Pfam; PF00560; LRR; C.
 DR Pfam; PF00561; NB-ARC; 1.
 DR Pfam; PF01582; TIR; 1.
 DR PRINTS; PRO0364; DISEASERSIST.
 DR SMART; SM00255; TIR; 1.

```

SQ SEQUENCE 1253 AA; 140809 MW; 4211ADE0566C2B01 CRC64;
Query Match 5.8%; Score 109.5; DB 10; Length 1253;
Best Local Similarity 20.3%; Pred. No. 1.2; Mismatches 130; Indels 127; Gaps 18;
Matches 81; Conservative 61;

Qy 77 VHLISLKNQRECS-----TMAVVSTPFFNRNLYWYLLSS-----RF 115
Db 421 IHL-----EGCGFFRVEINLVKCLVSM-EGRVVHNLQISGRKIINGKRRS 472

Qy 116 SLKSYCPFFLDYLEAFG---LUSDPLDHOAV-----IKFFEL-----ET 153
Db 473 RLWKPLIIRYFLEDRQVLSGSEIDAEIAPDPSALSFDVNPWAFENMYNLYLKICSSNFGN 532

Qy 154 HFSVYFVSFVAPHQVLSLQDRYPIASVMRTLDKDNFSL-----TPD-- 197
Db 533 HYALHLPKGVKSLPEELRLHWEHPPLSLPDNTRNLVILNMCYSKLQRLWEGTKELG 592

Qy 198 -LTHDLHGVPMHLHPSFSEFINN-----GRL-----FTKVB-----KVQ 233
Db 593 MLKRMILCHSQQLVGTQELQIALNEVIDLQGCARLQRFATGCHQHLRVINLSGCIK 652

Qy 234 ALPSKKORIOTL---OSNLIAIVRCFWFTVESGLIENHEGRKAYGAVGLISSPOELG-HAF 289
Db 653 SFVEVPNIEELVKGTGIRSIPTVTFSPQNSFIYDKDKHFLNREVSSDSQSLSIWY 712

Qy 290 IDNRVRLP-----LELDQIIRLPNT-----STQETLFSIRHPDELVEL-----TSKLEWM 336
Db 713 LDNLKVLDSLQCLELEDIOGIPRNRLKYLGGTAIKELPSLMLSLVLDLLENCKRLHK 772

Qy 337 LDQGL-----LESIPLYNQEKYLSG 356
Db 773 LPMGIGNLSLAVINLSGCGSELEDIOGIPRNLEELYLAG 811

RESULT 18
ID Q9GF63 PRELIMINARY; PRT; 506 AA.
AC Q9GF63;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Arabis alpina (Alpine rockcress).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabis.
OX NCBI_TaxID=50452;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AALP2;
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF144329; AAG43298.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MATK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01844; MATK_Nr.1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 506 AA; 60472 MW; 3BD5743D758B9F23 CRC64;

Query Match 5.6%; Score 106.5; DB 8; Length 506;
Best Local Similarity 22.6%; Pred. No. 0.73;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHSLSDRETFEPDPAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFEELVELTSKLEWLDQ 339
Db 62 VYFVSEFSFSAKEXKPTTYAGIKRPFVSRFPDPTNSVEVLDNPLKIQSGLEWVKDE 117

Query Match 5.6%; Score 106; DB 13; Length 129;
Best Local Similarity 25.0%; Pred. No. 0.16;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHSLSDRETFEPDPAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFEELVELTSKLEWLDQ 339
Db 62 VYFVSEFSFSAKEXKPTTYAGIKRPFVSRFPDPTNSVEVLDNPLKIQSGLEWVKDE 117

RESULT 20
Q9GF55
```

```

Matches 68; Conservative 37; Mismatches 131; Indels 65; Gaps 12;

Qy 3 YCE-RTLDPKYILKIALKRLQSLSPFQNSQSLQRAYSTPYSYRIILOKENKEKQALAR 61
Db 190 YCNWKQFD-----IKKKLILNRPFFLFLYNSHVC-----YESIFFLKRKSSHLRSTAY 239

Qy 62 HKCISILEFFKNL-----LFVH-----LLSLSKNQ-----REGCSTMAVVSTP----- 100
Db 240 EVFERILFYAKIQHFLKVFVNPFALGLGLKDPFLHYVRYHGKSLATKDTPLLMNKKW 299

Qy 101 FFRNLWYRLLSRFSMLKSYCPFFLDYLEAFGLISDFLDHOAVIKFFELETHFSYYPV 160
Db 300 FYFVNLWQYFVSQFQKIHINQISKDNLEFLGYLSSRLNPLVVRSQLNSFLI--- 356

Qy 161 SGFVAPHQVLSLQDRYPIASVMRTLDKDNFSLTPDLIHDLLGH-----VPLLHPSFSE 216
Db 357 -----DNIRIKLDNKIPISIIIGSLTKDKFC-----NLLGHPISKANN-TESSDS 401

Qy 217 PFNMGRLFTKVIKQVQALPSKKQR-----IOTLQSNLIAIVRCFWFTVESGLI 265
Db 402 ILNRFVIRCNISHYVSGSKKKHLYRIKYLRLCCVKTLARKHKSTVRAFLKRLGSLG 461

Qy 266 E 266
Db 462 E 462

RESULT 19
ID Q42428 PRELIMINARY; PRT; 129 AA.
AC Q42428;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Tyrosine hydroxylase (Fragment).
GN TH.
OS Lates calcarifer (Barramundi).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Centropomidae; Lates.
OX NCBI_TaxID=8187;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BARRAMUNDI;
RA Collet C., Candy J., Sara V.;
RT "Tyrosine hydroxylase and insulin-like growth factor-II but not
RT insulin are adjacent in the teleost species barramundi (Lates
RT calcarifer).";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007942; AAB64194.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioprotein_H; 1.
DR PRINTS; PR00372; FYHYDRXLASE.
FT NON TER
SQ SEQUENCE 129 AA; 14678 MW; 1AE29C7530AB4D78 CRC64;

Query Match 5.6%; Score 106; DB 13; Length 129;
Best Local Similarity 25.0%; Pred. No. 0.16;
Matches 29; Conservative 14; Mismatches 41; Indels 32; Gaps 1;

Qy 256 FWFVESGLTENHEGRKAYGAVGLISSPOELGHAFIDNVRLVPLLELDQIIRLPNTSTPOE 315
Db 2 YWFTVEYGLCKQNGEVKAYGAGLSSYGLVHSLSDRETFEPDPAAVQPYQDTYQP 61

Qy 316 TLF-----SIRHFEELVELTSKLEWLDQ 339
Db 62 VYFVSEFSFSAKEXKPTTYAGIKRPFVSRFPDPTNSVEVLDNPLKIQSGLEWVKDE 117
```

```

ID O9GF55 PRELIMINARY; PRT; 504 AA.
AC O9GF55;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Cardamine amara (Large bitter-cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Cardamine.
OK NCBI_TaxID=50461;
RN [1]
RP SEQUENCE FROM N.A.
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, BY
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
CC EMBL; AF144337; AAG43306.1; -.
DR Interpro; IPR000442; Intron_maturase2.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
SQ SEQUENCE 504 AA; 60332 MW; F557BF32927C7EF CRC64;

Query Match 5.6%; Score 106; DB 8; Length 504;
Best Local Similarity 22.8%; Pred. No. 0.81;
Matches 67; Conservative 36; Mismatches 115; Indels 76; Gaps 11;

OY 2 HYCERTPDPIYIKIALKQSLSPFONSQSLORAVSTPYSYRIILOKENEKQALAR 61
DB 214 HVCE-----YESIFFLKRSSHRSSTYEVLFERIFGYKIGHPLKVV 258
OY 62 HKCISILEFPKNLFFVHLISLKNORBEGCSTDMVAVSTP-----FPNNILYRLSSR 114
DB 250 -----IOHFFVFNPNPAILGLKDPFIHYVYHGRSLATKDPPLMNMKKYFVNLW 304
OY 108 YRLSSRPSLWKSYPFRFLDYLEAFGLSDPLDHOAVIKFELETHFSYPSGVAPH 167
DB 305 QCYFSVWFQSQVKHKLQSLKNLEFLGYLSRLNPLVRSQMLE-----NSFLIDN 356
OY 168 QYLSLQDRYPPIASVARTLKDQNFSLTPDLIHDLLGH---VPMILHPSFSEFFIMGR 223
DB 357 VRKLD--DSKIPISISIGSLAKKFC-----NVLGHPISKVTW--THSSSDILNPFVR 406
OY 224 LFTKVIKVKQALPSKQKOR-----IQTLSNLIAIVRCFWFVESGLIE 266
DB 407 ICRNISHYSSGSKKFLYRIKYILRLCCVKTLARKHKSIVRAFLKRLGSLLE 460

RESULT 21
O9GF31 PRELIMINARY; PRT; 504 AA.
AC O9GF31;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Cardamine penzancei.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Cardamine.
OK NCBI_TaxID=125588;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, BY
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
SQ SEQUENCE 504 AA; 60292 MW; 94C5F9B8282127A7 CRC64;

Query Match 5.6%; Score 106; DB 8; Length 504;
Best Local Similarity 22.3%; Pred. No. 0.81;
Matches 64; Conservative 41; Mismatches 120; Indels 62; Gaps 10;

OY 2 HYCERTPDPIYIKIALKQSLSPFONSQSLORAVSTPYSYRIILOKENEKQALAR 61
DB 214 HVCE-----YESIFFLKRSSHRSSTYEVLFERIFGYKIGHPLKVV 258
OY 62 HKCISILEFPKNLFFVHLISLKNORBEGCSTDMVAVSTP-----FPNNILYRLSSR 114
DB 259 NNPFALIGLKD--PFIHY-----RYHGRSLATKDPPLMNMKKYFVNLWQCYFSW 311
OY 115 FSLWKSYPFRFLDYLEAFGLSDPLDHOAVIKFELETHFSYPSGVAPHQYLSLQ 174
DB 312 FQSQVKHKLQSLKNLEFLGYLSRLNPLVRSQMLE-----NSFLIDNVRKLD-- 361
OY 175 DRYPIASVARTLKDQNFSLTPDLIHDLLGH---VPMILHPSFSEFFIMGRFLTKVTE 230
DB 362 DSKIPISISIGSLAKKFC-----NVLGHPISKVTW--THSSSDILNPFVRCRNISH 413
OY 231 KVQALPSKQKOR-----IQTLSNLIAIVRCFWFVESGLIE 266
DB 414 YSSGSSKKKLYRIKYILRLCCVKTLARKHKSIVRAFLKRLGSLLE 460

RESULT 22
O9GF30 PRELIMINARY; PRT; 504 AA.
AC O9GF30;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Cardamine rivularis.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Cardamine.
OK NCBI_TaxID=82338;
RN [1]
RP SEQUENCE FROM N.A.
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, BY
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
SQ EMBL; AF144365; AAG43334.1; -.

```

```
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 504 AA; 60302 MW; 66AA07B9E2C48C64 CRC64;

Query Match
Best Local Similarity 22.3%; Score 106; DB 8; Length 504;
Matches 64; Conservative 41; Mismatches 120; Indels 62; Gaps 10;

QY 2 HCYBRTDPKYLKIALKURQSLUFFONSOSLARAYSTPYVYRILQENKEKQALAR 61
DB 214 HVCB-----YESIFFLKRSLSRLSTSYEVLFERIFFYKIOHFLKVFV 258
QY 62 HXCISILEFFKNLLFVHLLSLSKNOREGCGSTDMVAVSTP-----FFNRNLWYRLSSR 114
DB 259 NNFPAILGLKLD-PFIHYV-----RYHGRSILATKDTPLLMNKKWYFVNLWQCVFSVW 311
QY 115 FSLWKSYCPREFLDYLEAGLLSDFLDHOAVIKFPELETHFSYVPVSGFVAPHOYLSLLQ 174
DB 312 FQSQKVHIKQLSKDNLEFLGYLSRLNPLVRSQMLE-----NSFLIDNVRIKL-- 361
QY 175 DRYPIASVMTLDKDNESLTDPDLIHDLGH-----VPMLLHPSPSEFFINMGLRFTKVIE 230
DB 362 DSKTPISTISGLAKKFC-----NVLGHPXSKVW-THSSDSDLNRFVRCINISH 413
QY 231 KVQALPSKKOR-----IQTLQSNLIAIVRCFWFTVESGLIE 266
DB 414 YYSGSKKKLYRIKYLRLCCVKTLARKHKSTVRAFLKRLGSGLLLE 460

RESULT 23
Q95F33 PRELIMINARY; PRT; 513 AA.
AC
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Muhlenbergia sylvatica.
OC Chloroplast
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chloridoideae; Eragrostidae; Muhlenbergia.
OX NCBI_TaxID=160557;
RN [1]
RP SEQUENCE FROM N.A.
RA Hilu K.W., Alice L.A.;
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF312350; AAK60046.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
FT NON_TER 513
SQ SEQUENCE 513 AA; 61341 MW; CD93D8582ECC3682 CRC64;

Query Match
Best Local Similarity 18.6%; Score 104; DB 8; Length 513;
Matches 67; Conservative 18; Mismatches 131; Indels 92; Gaps 16;

QY 22 QSLSLFFQNSQSLQR-----AYSTPYSYRILQENKEKQALARKHCISILE---FFKML 74
DB 199 KSIFLFSKDNKRLSRFLYNYFVSEYEFFLLRLKQS-----SCLRLTSSGTFLERI 249
QY 75 LFVHLLSLSKNOREGCGSTDMVAVSTPFFNRNLWY-----RLLSRFSL--- 117
DB 250 IFSRNE-----HFGVNYPGFWRIFWFMPLMHYVYQKAILASKGTLILK 298
QY 118 --WKSYPTRFELDYLEAGLLSDFLDHOAVIKFPELETHFSYVPVSGFVAPHOYLS---L 172
DB 299 KWKSKYLWNVSQVFLFTQPORICLNQLTNSCDFLGYSRVPMQNFVLRNQMLENSEFL 358
QY 173 LQ-----DRYFFIASVMTLDKDNFSLTDPDLIHDLGH-VPMLLHPSPSEFFI--NMGR 223
DB 199 KSIFLFSKDNKRLSRFLYNYFVSEYEFFLLRLKQS-----SCLRLTSSGTFLERI 249
```

```

Db 359 ITRAKKPTTVPVTSIGLSKAKFCTG-----LGHPIKPIWTDLSDMDILDRGR 411
Qy 224 LFTKVIKVOALPSKKQRIQTLQSNLAIYRCFMFVSGLENNHGRKAYGAVLISSPQ 283
Db 412 ICRNLPH-YHSGSSKKQLYRLK---YILR---LSCARTLARKH-----STVRFMQ 457
Qy 284 ELGHAIFIDNVRVLPLELDQIIRLPFNTSTPOETLFSIRHPELVELVTSKLEWMLD 338
Db 458 RLGSVFLBEFFT---EEQVFSIMFTKT-----HF-SFGHLSERIWYLD 499

RESULT 25
ID 095F39 PRELIMINARY; PRT; 513 AA.
AC 095F39;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Monanthochloe litoralis.
OC Chloroplaet.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chlorididae; Eragrostidae; Monanthochloe.
ON NCBI_TaxID=160556;
RN 11
RP SEQUENCE FROM N.A.
RA Hilu K.W., Alice L.A.;
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF312349; AAK60040.1; -
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KM Chloroplaet; mRNA processing.
SQ SEQUENCE 513 AA; 61512 MW; FA97486C266572D8 CRC64;

Query Match 5.5%; Score 103.5; DB 8; Length 513;
Best Local Similarity 21.0%; Pred. No. 1.3; Indels 89; Gaps 19;
Matches 75; Conservative 59; Mismatches 134;

Qy 12 YLTKIALKRLQSLIFPONSQSLQRAYSTPYSYVRIILQENKKEKQALARRKICISILEFF 71
Db 202 FLK---KENRRLSRFLYNS-----YVSEYEFLLPLRKOS-----SCLRLA 240
Qy 72 KLLLFVHLLSLKNQRECGSDMAVSTPFENLWY-----RLSSRPSL 117
Db 241 SSGTFLERIHFSRKMEH-----FCWYRGFFRTIWFMDPLMHVYRQGYLLASKCTL 295
Qy 118 -----WKSYPREFLDYLEAAGLSDFLDHQAVIKFPELETHSYVPYSGFAAPROYLSL 172
Db 296 LFGKKMKSYLVNFSQYFSPFTQQRRLANQLNSCFDILGRSSVPIINTPLVANKML-- 353
Qy 173 LODRYFPFIASVMRLDKDNSTLPDLIHD-----LGH-VPMLLHPSSEFT--NM 221
Db 354 --ESFFPFIATMKKPD--TPAPAP--LIGLSKQOCTGAGHPISKPIWTDLSDMDILDRF 409
Qy 222 GLFTKVIKVOALPSKKQRIQTLQSNLAIYRCFMFVSGLENNHGRKAYGAVLISS 281
Db 410 GICCNLPH-YHSGSSKKQLYRLK---YILR---LSCARTLARKH-----STVRF 455
Qy 282 POLGHAIFIDNVRVLPLELDQIIRLPFNTSTPOETLFSIRHPELVELVTSKLEWMLD 338
Db 456 MÖRLGSVFLBEFFT---EEQVFSIMFAKTT-----HFSFGHSERI-WYLD 499

```

```

RESULT 26
ID 09VP19 PRELIMINARY; PRT; 1782 AA.
AC 09VP19;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CG11376 protein.
DE CG11376.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10711132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abbil J.F., Ashbyant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari P., Brotier P.,
RA Burkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Buttle K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodde C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goeler A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegian C.,
RA Jallai B., Kalush F., Karpen G.H., Ke, Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matcel B., McInosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M.,
RA Palazzolo K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinet K., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Shue B.C., Spadling A.C., Stapleton M., Strong R., Sun E.,
RA Spter E., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Maasarm D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhao Q., Zheng L.,
RA Zhibg X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AB003590; AAF51561.1; -
DR FLYBASE; FBgn003216; CG11376.
SQ SEQUENCE 1782 AA; 202676 MW; 3C80FF5EC1705E1B CRC64;

Query Match 5.5%; Score 103.5; DB 5; Length 1782;
Best Local Similarity 25.0%; Pred. No. 6.1; Indels 91; Gaps 15;
Matches 73; Conservative 32; Mismatches 96;

Qy 67 ILRF-----FKNLLFVHLLSLKNQRECGSDMAV-----VSTPFENRLWYR 109
Db 574 ILRFPOSALYNPHYSYRNLFFSPPELNFSSRASANNIAVRVQLMAGSTPDAAVAAY- 632
Qy 110 LLSRSFIMKSCYPRFFLDYLEAFGLS-----DFPDHDAVYKFELETH-----FSYYP 159
Db 633 -----GKSSCPK---STEAFAVAVYHNKCSFDFEIKIALPASIKOHHLHFTIYH 681
Qy 160 VSGFVAP-----HGYLSLDQRYRPIAS-----VMRTLDKDNSTLPDLIHDLL 203

```

```

Db 682 VSCQKQDQPSVETPIGYTLPLEDEGKLFGEFNLPMVVFSPPENYSFIPPNVH--L 739
Qy 204 GHVPLLLHPSFSEFFINMGRLETKVIEKQVALPSKQRIQTLOSNIIA-IVRCFWFTVES 262
Db 740 PGIKWL-----DNHRAVESINVEAVTA-----IHTLSDFDRFFLICEYLDTRN 783
Qy 263 GLIENHGRKAYGAVLISSPQELCHAFID-----NVRVLPLELDQIIRL 306
Db 784 --IPSHIGEN-----NIETELKCLLDIEYANREPLVRHLPIVLDKLIEL 826

```

RESULT 27

```

Q9GFA1 ID Q9GFA1 PRELIMINARY; PRT; 504 AA.
AC Q9GFA1; 2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Aubrieta.
OX NCBI_TaxID=81984;
RN [1]
RP SEQUENCE FROM N.A.
RA Koch M., Mitchell-Olds T.;
RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
RT synthase and their utility for phylogenetic reconstructions within the
RT Brassicaceae."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF144352; AMG43321.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 504 AA; 60023 MW; 4620F813810B3AD8 CRC64;

```

```

Query Match 5.5%; Score 103; DB 8; Length 504;
Best Local Similarity 21.9%; Pred. No. 1.5;
Matches 66; Conservative 37; Mismatches 113; Indels 86; Gaps 12;

Qy 8 LDPKYILKIALKRLQSLSPFONSQSORAYSTPYVYRILOKENKEKQALARKHCISI 67
Db 202 LNPRFL-----LFYLNHVC-----YESIFFLR-----KRSSPLRSTAYEV 239
Qy 68 LEFFKNLFLV-----HLLSLKQ-----REGCTDMNVSTP-----100
Db 240 L--FERILFYGIHFLKLVFNFLTPGLLKDPLHVRVHGKSLIATKDTPLMNKWK 297
Qy 101 FPNRLWRLSSRFLSWKSCYCPRFELDYLEAFGLSDPLDQHVAKVFFLETHFSYYPV 160
Db 298 FVFVNLWCQYPSVMFQSKYNINQLSKNLEFLGYLSRLNPLVRSQMLENSFLI--- 354
Qy 161 SGFVAPHQYLSLQDRYFFPIASVMRTLDKDNFSLTPDLIHDLLGH-----VPVLLHPSFE 216
Db 355 -----DNIRIKLDSKIPISSTMGSLAKDKFC-----NVLGHPISKATWTDSDDFD- 399
Qy 217 PFINMGRFLTKVIEKQVALPSKQRIQTLOSNIIAIVRCFWFTVESGL 264
Db 400 -ILNRFVRCIRNISHYSGSKKNLYRIKYLRLCCVCKTLARKHKSTVRAFLKRVGSL 458
Qy 265 IE 266

```

```

Db 459 VE 460
RESULT 28
Q9TIB8 ID Q9TIB8 PRELIMINARY; PRT; 513 AA.
AC Q9TIB8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Bouteloua curtipendula.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chloridoideae; Cynodonteae; Bouteloua.
OX NCBI_TaxID=110875;
RN [1]
RP SEQUENCE FROM N.A.
RA Hilu K.W., Alice L.A.;
RT "Phylogenetic relationships in subfamily Chloridoideae (Poaceae) based
RT on matk sequences: A preliminary assessment."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL; AF144578; AAF20334.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 513 AA; 61519 MW; 223EBA9CEA96DFEB CRC64;

Query Match 5.5%; Score 103; DB 8; Length 513;
Best Local Similarity 22.0%; Pred. No. 1.5;
Matches 78; Conservative 59; Mismatches 127; Indels 90; Gaps 22;

Qy 22 QSLSLFQNSQSOR-----AYSTPYVYRILOKENKEKQALARKHCISILEFFKNLLFV 77
Db 199 KSIFLFSKENKRLSRFLYNSYVSEYEFLLFLRKQS-----SCLRTSSGTFR 246
Qy 78 HLLSLKNQREGCSTDMNVSTPFFNRNLW-----YR---LLSSRFSL-----W 118
Db 247 ERIFFSRMEH-----FGVMYPGFELKTIWIFMDPLMHVYRQGVILASKGTLFLFQKKW 301
Qy 119 KSYCPRFELDYLAFAGLSDFLD-HQAVIKFFELETHFSYYPVSGFVAPHQYLSLQDRY 177
Db 302 KSYLVN-FSQYFFSFWIQPQIRINQLTNSCFDGLGVHSSVPINTFLVRNQL-----ENF 356
Qy 178 PFIASVMTKDKNFSLTDLIHDL-----LGH-VPVLLHPSFSEFFI--NMGRFLT 226
Db 357 FLIATRMKKFD-TTVATP-LTGSLSKAQFCCTGLGHPISKPIWTDLSLDWILDRFGICR 414
Qy 227 KVIEKQVALPSKQRIQTLOSNIIAIVRCFWFTVESGLIENHGRKAYGAVLISSPOELG 286
Db 415 NLFH-YHSGGSKQTYLRUK-----YILR---LSCARTLARKHK-----STVTFMQRLG 460
Qy 287 HAFIDNVRVLPLELDQIIRLPFNVTSTPQETLFSIR--HFDDELVELTSKLEWMLD 338
Db 461 SVFLEEFFT---EEQVFSLMF-----PRATHFSFHGSHSERI-----WYLD 499

RESULT 29
Q95F60 ID Q95F60 PRELIMINARY; PRT; 513 AA.
AC Q95F60;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

```


DE Probable intron maturase (Maturase K).
GN MATK
OS Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chlorididae; Chloridace; Buchloe.
OX NCBI_TaxID=160577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TEXOKA;
RA Hilu K.W., Alice L.A.;
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AF312325; AAK60016.1; -
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
KW SEQUENCE 513 AA; 61509 MW; 1162PCCS17CADDID CRC64;
SQ
Query Match 5.5%; Score 103; DB 8; Length 513;
Best Local Similarity 21.5%; Pred. No. 1.5; Indels 90; Gaps 21;
Matches 76; Conservative 59; Mismatches 129; Indels 90; Gaps 21;
QY 22 OSLSLFFQNSOSIOR---AYSTPYRYRIILOKENEKQALAHKICISLFFQNLFFV 77
DB 199 KSIFLFSKENKRLSRFLYNYSVSEYEFLLFLRKOS-----SCLMLTSSGTF 246
QY 78 HLILSLKQRECGSTDMAVVSTPFNNLMW-----YR---LSSRPSL-----W 118
DB 247 ERHIFSRKMEH-----FCVMYRGFFRKTIVFMQPLMHVRYQGIILASKQTLDFQKW 301
QY 119 KSYCPFFLDYLEAFGLSDFLD-HQAVIKFELETHFSYYPVSGFVAPHQYLSLDORY 177
DB 302 KSYLVNFSQYFSPMAQROKICINQITNSCDFLGYSSVINTFVLTQML-----ENF 357
QY 178 FPLASVMTLDKDNFSLTPLDILHDL-----LGH---VPMILHSFSEFFINMGRLEFT 226
DB 357 FLATRMKKFD-TTVPTVP-LIGSLSKAQFCTGLGHPISKPIWTDLSDGDIIDRGRICR 414
QY 227 KYIEKVALPSKKORIQTQSNLIAIVRCMFVSEGLIENHGRKAYGAVLISSPQELG 286
DB 415 NLFH-YHSGSSKKRTLYRLK---YLIR---LSCARTLARKK-----STVTFMQRLG 460
QY 287 HAFIDNVRLPLELDQIIRLPNTSTPOETLFSIR-HPEDELVELTSKLEMLD 338
DB 461 SVFLSEFFY---EEQVFLYLMF---PKTHFSFHGSHSRI-----WYLD 499
RESULT 30
Q95F50 PRELIMINARY; PRT; 513 AA.
ID 095F50
AC 095F50
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Schedonardus paniculatus.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chlorididae; Chloridace; Schedonardus.
OX NCBI_TaxID=160583;
RN [1]
RP SEQUENCE FROM N.A.

RA Hilu K.W., Alice L.A.;
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AF312325; AAK60026.1; -
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
KW SEQUENCE 513 AA; 61384 MW; 94861P9870BHF8BA CRC64;
SQ
Query Match 5.5%; Score 103; DB 8; Length 513;
Best Local Similarity 18.6%; Pred. No. 1.5; Indels 102; Gaps 16;
Matches 67; Conservative 64; Mismatches 127; Indels 102; Gaps 16;
QY 22 OSLSLFFQNSOSIOR---AYSTPYRYRIILOKENEKQALAHKICISLFFQNLFFV 77
DB 199 KSIFLFSKENKRLSRFLYNYSVSEYEFLLFLRKOS-----SCLMLTSSGTF 246
QY 78 HLILSLKQRECGSTDMAVVSTPFNNLMW-----YR---LSSRPSL-----W 118
DB 256 -----EHFGVMYRGFFRKTIVFMQPLMHVRYQGIILASKQTLDFQKW 301
QY 119 KSYCPFFLDYLEAFGLSDFLD-HQAVIKFELETHFSYYPVSGFVAPHQYLSLDORY 178
DB 302 KSYLVNFSQYFSPMAQROKICINQITNSCDFLGYSSVINTFVLTQML-----ENF 357
QY 179 FPLASVMTLDKDNFSLTPLDILHDLGHVPMILHSFSEFFINMGRLEFTK-----V 228
DB 358 LIATRMK-----KFDITTPPAT-PLIGSL-----SKQFCGSGHPISKPIWTDLSDMDI 405
QY 229 IEKVALP-----PSKKORIQTQSNLIAIVRCMFVSEGLIENHGRKAYGAVLI 278
DB 406 LDRFGRICNMLFHYHSGSSKKQTLVRLK---YLIR---LSCARTLARKK-----STV 452
QY 279 ISSPQELGHAFINNVRLPLELDQIIRLPNTSTPOETLFSIR-HPEDELVELTSKLEMLD 338
DB 453 RTFMQRLGVSFLSEFFY---EEQVFLYLMF---PKTHFSFHGSHSRI-----WYLD 499
RESULT 31
Q95F38 PRELIMINARY; PRT; 513 AA.
ID 095F38
AC 095F38
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Reederochloa eludens.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chlorididae; Eragrostideae; Reederochloa.
OX NCBI_TaxID=160560;
RN [1]
RP SEQUENCE FROM N.A.
RC "A Phylogeny of Chloridoideae (Poaceae) Based on matK Sequences."
RT Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS (BY SIMILARITY).
CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
CC MITOCHONDRIAL INTRONS.
DR EMBL: AF312350; AAK60041.1; -
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.

RA Pfam: PF01348; Intron_maturae2; 1.
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RT EMBL: AE003617; AAF52504.2; -;
DR HSSP: P08799; 1MND.
DR FlyBase: FBgn002938; ninaC.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR001245; Ty_pkinase.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000001; Euk_pkinase; 1.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 2.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
DR ATP-binding; Transferase.
KW SEQUENCE 1154 AA; 134334 MW; 211CD4A2F295B9A2 CRC64;
Query Match 5.4%; Score 102.5; DB 5; Length 1154;
Best Local Similarity 22.7%; Pred. No. 4.4;
Matches 82; Conservative 56; Mismatches 129; Indels 95; Gaps 20;
QY 10 PRYLIKIALKRLSLSIFFNSQSLQRAVSTPYSYRILIQENKEKQALARKH-----CI 65
DB 839 PEMIETFRSSLDSEIMLFMTNQLTKAGNLTPW---FEAVQHKESERKSYALNTLSAGCI 895
QY 66 SILPEFFKNLLFVLLSLSKNREGCGTDVAVSTPFFNRNLWYRLSS--RFLSWKSYCPR 124
DB 896 SQVN-----NRLTLAANFRFTCLTLKMLSQ--NANLGVHFVRCIRADL--EYKPR 942
QY 125 FF-----LDYLEAFGLSDFDLHOAVIKFFETHFSYPSVSGFVAPHVLSLLODRYPP 179
DB 943 SFHSDVVQQMKALGVLDTVIARQ---KGFS-----SELPDFELRYQLAFDFDE--P 992
QY 180 TASVWRLTDKNFSLTPDLIHDL-----LGHVPMLLHPSPSEFFINMGELFTKVIKQV 233
DB 993 V-----EMTKCNEL---LFLKMEGHALGKTKVFLRYNDF---LAEVLQVKV- 1040
QY 234 ALPSKKQRIQTLOSNLIAIVRCFWFTVSGLI-----ENHEGRKAVGAVLISPBQL 285
DB 1041 -----IKVQSMRWALLARKR-----VKGKGVFLKGGKGGEHHD-----VAASKIQKA 1082
QY 286 GHAFIDNVRLVPELDOILRL-----PNTSTPQETLFSI-----RHFDLVELVT 330
DB 1093 FRGFDRFDRVLPPLVNEKSGQLNENTADTFRPFAKKWREKSIFQVLLHYRAARFQDFVNL 1142
QY 331 SK 332
DB 1143 QQ 1144
RESULT 33
ID Q9GFM4 PRELIMINARY; PRT; 506 AA.
AC Q9GFM4

DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Probable intron maturase (Maturase K).
 GN MATK.
 OS Arabis alpina (Alpine rockcress).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabis.
 ON NCBI_TaxID=50452;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=ALP1.
 RA Koch M., Mitchell-Olds T.;
 RT "Evolutionary analysis of plastidic maturase K and nuclear chalcone
 RT synthase and their utility for phylogenetic reconstructions within the
 RT Brassicaceae.";
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS, BY
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 CC EMBL; AF144328; AAG3297.1;
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR02866; MatK_N.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 DR Chloroplast; mRNA processing.
 KW SEQUENCE 506 AA; 60662 MW; 275A64406CE9525 CRC64;
 SQ
 Query Match 5.4%; Score 102; DB 8; Length 506;
 Best Local Similarity 22.9%; Pred. No. 1.8;
 Matches 67; Conservative 35; Mismatches 117; Indels 74; Gaps 11;
 QY 2 HYCERTLPKYLTKALRKOSLSLFONOSLOPASTPYSYVYIIQKNEKQOLAR 61
 DB 216 HYCE-----YESIFFLARKSSHLRSTAEVLEFRLF-----YAKIQ 253
 QY 62 HNCISLEFKNLFLVHLISLKNQ-----REGGCTDMAVSTP-----FNNRNLMYR 109
 DB 254 H---FLKVFYV-NFPAIIGLLKDPFLHYVYHGXSLTKDTPILMKKWFYVNLMOF 308
 QY 110 LLSRFSLSKSYCPREFLDYLEAFGLISDFLDHQAIVKFELETHPSYVSGVAPHOY 169
 DB 309 YFSVWFQSKTHINOLSKDNLEFLGYLSLRNPLVVRQMLENSFLI-----DN 358
 QY 170 LSLQDRYFPPIASVWRTLDKDNFSLTPDLIHDLGH---VPMLLHPSFSEFFINMGRLF 225
 DB 359 IRIKLDNKIPISISIIIGSLTKDKFC-----NLGHPISKAWMT--ESSDSIDINRFLRI 409
 QY 226 TVYIEKVALPSKKQR-----IOTLOSNIATVRCFVFESGLIE 266
 DB 410 CNHISHYSSGSKKKHLVRIKYLRLCCVKTLLARKKSTVRAFLKRLSGGLE 462
 RESULT 34
 ID 09MV23 PRELIMINARY; PRT; 519 AA.
 AC 09MV23;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Probable intron maturase (Maturase K) (Fragment).
 GN MATK.
 OS Oryza meyeriana.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzaceae; Oryza.
 ON NCBI_TaxID=83307;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=20056256; PubMed=10588717;
 RA Ge S., Sang T., Lu B.R., Hong D.Y.;
 RT "Phylogeny of rice genomes with emphasis on origins of allotetraploid
 RT species.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:14400-14405(1999).
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 CC EMBL; AF148673; AAF3181.1;
 DR InterPro; IPR000442; Intron_maturase2.
 DR InterPro; IPR02866; MatK_N.
 DR Pfam; PF01348; Intron_maturase2; 1.
 DR Pfam; PF01824; MatK_N; 1.
 DR Chloroplast; mRNA processing.
 KW NON_TER
 FT NON_TER 519 519 1
 SQ SEQUENCE 519 AA; 62257 MW; 0148830573EB14C1 CRC64;
 Query Match 5.4%; Score 101.5; DB 8; Length 519;
 Best Local Similarity 24.5%; Pred. No. 2;
 Matches 96; Conservative 55; Mismatches 148; Indels 93; Gaps 26;
 QY 1 VHYCERTLPKYLTKALRKQ-----SLSL--PFONOSLOPASTPYSYVYIIQ 50
 DB 165 LHYLSHIEIYPHIDLILOLQYRIQVPSHLRPFLLNYSSMNSFITSMKSI-LLIK 223
 QY 51 KENNEKOLARHKCISILEFFKNLFLVH---LISLKN---QREGGCTDM---AVYST 99
 DB 224 KENRRLPRLNYSYSEYEFF--LFLRKQSGCLRLTSSGTPLEIRHFRKKHNGVWP 281
 QY 100 PFNNRNLMY-----RLISGRFL-----WKSVCREFLDYLEAFGLISDPL 140
 DB 282 AFRKTVYFPMDDIMHYVYQKAILASKGTLLKKKCYLYRLM-QYSPSFWTQSRRI 340
 QY 141 D-HQAVIKFPELETHSYSPVSGVAPHOYISLQDYFPPIASVWRTLDKDNFSLTPDL 199
 DB 341 HLNOLANSCTDFLGTSVYVINSLSLVKNQML---ENSLFDTRMKFPD-TVYAPRP-LI 394
 QY 200 HDLL-----GH-VPMLLHPSFSEFT--NNGRLFTKVIKVALPSKKRIOTLOS 248
 DB 395 GSLAKAQFTGSGHPISKPRTWDSWDIDLDFRIGRNLFH-YHSGSKKKTLYRLK-- 451
 QY 249 LIAIVRCFVFESGLIENHEGRKAYGAVLISPOELGNAFIDNVVRLPLELDQIIRLPF 308
 DB 452 --YILR--LSCARTLARKHK-----STVRAFPWOMLGSVFLREFT--EEEOVFSIMF 497
 QY 309 NTSTPQETLSIR--HFDLVELTSKLEWMD 338
 DB 498 ---AKTYFFSFGSHSDRI-----WYLD 517
 RESULT 35
 ID 08T2D4 PRELIMINARY; PRT; 1615 AA.
 AC 08T2D4;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypoetical 192.5 kDa protein.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 ON NCBI_TaxID=44689;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=AX4.
 RA Gloeckner G., Eichinger L., Szafarski K., Pachebat J., Dear P.,
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
 RA Tunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.

Query Match 5.3%; Score 101; DB 17; Length 635;
Best Local Similarity 19.8%; Pred. No. 2.9;
Matches 79; Conservative 69; Mismatches 123; Indels 128; Gaps 24;

Qy 25 SLFFPNQSQR-----AVSTPSYVRIILQENKE-KQALAHKCTSIILEPFK 72
 ||||| : ::||| |::| : :
 : : : : : :

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.H., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003462; AAF47145.1; -;
 DR FLYBASE: FBgn0034964; CG3173.
 SQ SEQUENCE 2042 AA; 233746 MW; 9144DB7B51F52BES CRC64;

Query Match 5.3%; Score 101; DB 5; Length 2042;
 Best Local Similarity 21.1%; Pred. No. 12;
 Matches 78; Conservative 49; Mismatches 122; Indels 120; Gaps 19;

OY 48 ILKENKEKQALAHKCSILEPKNL-----LFFHLISLNKREGC 90
 DB 879 IINENSRLDEL-----VNFILRNIFDGNLSHQIVGELLDYIFRLSTVKQSVAA 930
 OY 91 STDMAVY---STPFNNLWYRLSS-----RFSLMKSCYCPFLDY 129
 DB 931 LSGKTIIRHSQDF--ENEM--LKSLOQIRHYEVKPIIRPOLAAQVENCCELLMAX 966
 OY 130 LE--AFGL--LSDFLDH-----QAVIKFELETH-----FSYFVSG-----F 163
 DB 987 IOFTTAHTLNDPVNEMLDHVIDMAQGLIVERSTWFOHIIISQSDYIVFDENRITQLKLP 1046
 OY 164 VAPHOYLSLQDRYFPPIASVWRITLDKDNFSLPDLI--HDLLGHVPLWLHPSESEFFIN 220
 DB 1047 VMFNYYIILKLEHYEPY-----EMTEYFDLIMVQFDGVOLP--LH-----IN 1087
 OY 221 MGRLEFTVIEKVALPSKRIQTLQSNLAIIVRCFWPTVESGLIENHGRKAYGAVILIS 280
 DB 1088 ITHAF--IILTYSNSNMPSIPILD-----YMPF-----PGRPAVAFAIPS 1127
 OY 281 SFQELGHAFFID--NVAVLPLELDQIIRLPNTSTPOETLFSIRHDELVEYTSKLEWMLD 338
 DB 1128 MPQEOVQLPMLKLMIRSSVDRLEFALINDLTPQIVLFVQNGFTVNSMSKILMLD 1187
 OY 339 OGLESIP, 347
 DB 1168 TAVLEQFPL 1196

RESULT 38

ID P97517 PRELIMINARY; PRT; 102 AA.

AC P97517, P70468;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Tyrosine hydroxylase (fragment).
 OS *Phodopus sungorus* (Striped hairy-footed hamster) (Djungarian hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC *Phodopus*.
 OC NCBI_TaxID=10044;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HYPOPHALAMUS;
 RA Bockman J.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y09294; CAA70476.1; -;
 DR HSSP: P04177; ITOH.
 DR InterPro: IPR001273; Aaa_hydroxylase.
 DR Pfam: PF00351; biotinyl_H_1.
 DR PRINTS: PR00372; FFMHYDKLASE.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 102 AA; 11433 MW; 52241A9D5DB27795 CRC64;

Query Match 5.3%; Score 100.5; DB 11; Length 102;
 Best Local Similarity 29.3%; Pred. No. 0.35; Indels 7; Gaps 1;
 Matches 27; Conservative 15; Mismatches 43;

OY 234 ALPSKRIQTLQSNLAIIVRCFWPTVESGLIENHGRKAYGAVILISFQELGHAFFIDNV 293
 DB 10 SLGASDELKST-----YMFTEFGKONGELKAYGAGLLSYGELHLSSEEP 62
 OY 294 RVLPLELDQIIRLPNTSTPOETLFSIRHDE 325
 DB 63 EVAFDPEAAMQPVODQTYQPVYVSSFSF 94

RESULT 39

ID Q9D0D1 PRELIMINARY; PRT; 2212 AA.

AC Q9D0D1
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE L protein.
 GN L.
 OS Ebola virus (strain Zaire Mayinga) (Ebo).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
 OC Ebola-like viruses.
 OC NCBI_TaxID=128952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAYINGA;
 RC MEDLINE=99171703; PubMed=10073695;
 RA Volchikov V.E., Volchikova V.A., Chepurnov A.A., Blinov V.M., Dolnik O.,
 RA Netesov S.V., Feldmann H.;
 RT "Characterization of the L gene and 5' trailer region of Ebola
 RT virus.";
 RL J. Gen. Virol. 80:355-362(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAYINGA;
 RC PubMed=11062045;
 RA Volchikov V.E., Chepurnov A.A., Volchikova V.A., Ternovoj V.A.,
 RA Kleen H.D.;
 RT "Molecular Characterization of Guinea Pig-Adapted Variants of Ebola
 RT Virus.";
 RL Virology 277:147-155(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAYINGA;
 RA Volchikov V.E.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF272201; X84017.1; -;
 DR InterPro: IPR002877; Ftsd.
 DR InterPro: IPR001016; Viral_RNA_pol_L.
 DR Pfam: PF00946; Paramyx_RNA_pol_1.
 DR SEQNAME: 2212 AA; 252692 MW; 5B34COA291603DAB CRC64;

SO

Query Match 5.3%; Score 100.5; DB 12; Length 2212;
 Best Local Similarity 23.1%; Pred. No. 14;
 Matches 62; Conservative 42; Mismatches 95; Indels 69; Gaps 13;

OY 13 ILKIALKRLQSLFPPNSQSLQRAVSTPYRYRIILQKENKEKQALAHKCSILEPFR 72
 DB 355 LIRLEMPQOLCELFF-----SIQKWHGHPVLHSETAIQYVKATVLLKALRPVIVETV- 408
 OY 73 NLFFVHLISLNKQREGCTDMAVVS-----TFNNLWYRLSSRFSLMKSCYCPR 124
 DB 409 ---CVFKYSIAKIVPDSQGSWYSVSDRNLTPLGNSYIKKNQPPPLMKELMEFY--- 462
 OY 125 FPLDYLEARGL-----LSDFLDHQAVI-----KPELEHNPYSYVSGFVAPHOYSL 172
 DB 463 -HDDHPPLFTKTIISLIFIKQATVAVERTCMQAVDEPN-----VLQYNPFRKSTK 514

QY 173 LODRYFFPIASVMTLKDKNFSLTPDLIHLGHVPL--LHPSFSEFF-----INMR 223
DB 515 RVPEQF-----LEQENS-----IENVLSYAQKLYLLPQYRNFSLKEKELNVR 561
QY 224 LFTKVIKQVALPSKKQRIQTLQSNLIA 251
DB 562 TFGK-----LPYPTNRVQTLCEALLA 582

RESULT 40
Q9TI97
ID Q9TI97 PRELIMINARY; PRT; 513 AA.
AC Q9TI97;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Sporobolus indicus (smut grass).
OC Chloroplast
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chloridoideae; Eragrostidae; Sporobolus.
OX NCBI_TaxID=38731;
RN [1]
RP SEQUENCE FROM N.A.
RA Hilu K.W., Alice L.A.;
RT "Phylogenetic relationships in subfamily Chloridoideae (Poaceae) based
on matk sequences: A preliminary assessment."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS.
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
MITOCHONDRIAL INTRONS.
DR EMBL: AF144601; AAF20357.1; -;
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 513 AA; 61561 MW; BF6BB16048B3ECC6A CRC64;

Query Match 5.3%; Score 99.5; DB 8; Length 513;
Best Local Similarity 23.3%; Pred. No. 3;
Matches 91; Conservative 54; Mismatches 153; Indels 93; Gaps 20;
QY 5 ERTLPKYLKALK-----LRQSLFFQNSQSQ--RAYSTPYSYR-----46
DB 142 EKFLHLVLSHIEPIYPHPEILLVQLLEYRIQDVESLHLLRFFLLNYYSNNLSITSMKSI 201
QY 47 ILOKENEKQALARKHCISILEFFKNLLFVHLLSLKN-----QREGCSTDN--A 95
DB 202 FLKKENKRLPRFLYNSVVEYEFF--LLFLRKQSSSLRLISSGTELELHFHSMKMEHFG 259
QY 96 VVSTPFFNRNLWY-----RLLSRFSL-----WKSVCPRFLDYLEAFGLL 136
DB 260 VMYPRFQKTLWPMFMDPLMHVRYQKAILASKGTLLKXKXSYLVNFSQYFLSPWTP 319
QY 137 SDFLDHQAVIKFELETHPSYVPVSGVAPHQYLSLLQDRYFPIASVMTLKDKNFSLTP 196
DB 320 QRILNQLNRNSCDFLYGRSVNPIVFNQML-----ENFLDTRKKLD--TTAPATP 374
QY 197 DLHLDL-----LGH-VWMLHPSFSEFFI--NWRLFTKVKTEKQVALPSKKQRIQTL 245
DB 375 -LIGSLSKAQCTGLGHPISKPTLWTDLSWDILDRFGICRNLFH-YHSGSSKKQALYOL 432
QY 246 QSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLELDQIR 305
DB 433 K-----YILR-----LSCARTLARKH-----STVTFMQLRGLSVLEEFFFT---EEQVFS 476
QY 306 LPFNTST-----PQETLFSIRHFDLV 327

DB 477 LMFAKTTHFSFGSHSERIWFYDIIRIDDLV 507
RESULT 41
Q95F61
ID Q95F61 PRELIMINARY; PRT; 513 AA.
AC Q95F61;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK
OS Aegopogon cenchroides.
OC Chloroplast
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Chloridoideae; Chlorideae; Aegopogon.
OX NCBI_TaxID=160573;
RN [1]
RP SEQUENCE FROM N.A.
RA Hilu K.W., Alice L.A.;
RT "A Phylogeny of Chloridoideae (Poaceae) Based on matk Sequences."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
INTRONS (BY SIMILARITY).
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
MITOCHONDRIAL INTRONS.
DR EMBL: AF312324; AAK60015.1; -;
DR InterPro: IPR000442; Intron_maturase2.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 513 AA; 61640 MW; E80743D3EA97F086 CRC64;
Query Match 5.2%; Score 99; DB 8; Length 513;
Best Local Similarity 19.1%; Pred. No. 3.3;
Matches 69; Conservative 63; Mismatches 124; Indels 106; Gaps 18;
QY 22 QSLSLFFQNSQSQ--AYSTPYSYRILLOKENEKQALARKHCISILEFFKNLLFV 77
DB 199 KSIFLPSKRNKLSRFLYNYFVSEYEFFLLRQKS-----SCLRTSYGTFLERI 249
QY 78 HLLSLKNQREGCSTDMAVSTPFFNRNLWY-----LLSRFSL-----W 118
DB 250 HFFRMEH-----FGVNYFQFQKTIWFLMELMHVRYQKVLASKGTLLPQKKM 301
QY 119 KSYCPRFLDYLEAFGLLSDFLDQHAVIKFELETHPSYVPVSGVAPHQYLSLLQDRYF 178
DB 302 KSYVNFESQYFFFFMAQPORICLNQUTNSCDFLYGRSVNPIVFNQML-----ENFF 357
QY 179 PIASVMTLKDKNFSLTPDLIHLGHVPLHPSFSEFFINMGRFLT-----V 228
DB 358 LIATRMKKFD--TTAPATP-LIRSL-----SKAQFCTGSGHPISKPIWTDLSWDI 405
QY 229 IETKQAL-----PSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
DB 406 LDRFGICRNLPHYSGSSKKQTLVRLK-----YILR-----LSCARTLARKH 452
QY 279 ISSPQELGHAFIDNVRVLPLELDQIRIPFNTSTPQETLSIR--HFDLVELTSLKLEW 336
DB 453 RTFMRQLRGLSVLEEFFFT---EEQVFSLMFS-----KITYFSFHGSHSERI 497
QY 337 LD 338
DB 498 LD 499
RESULT 42
Q95F52
ID Q95F52 PRELIMINARY; PRT; 513 AA.
AC Q95F52;

DT 01-DEC-2001 (TREMELREL. 19, Created)
 DT 01-DEC-2001 (TREMELREL. 19, Last sequence update)
 DT 01-MAR-2002 (TREMELREL. 20, Last annotation update)
 DE Probable intron maturase (maturase K).
 GN MATK.
 OS Gymnopus brevifolius.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Chloridoideae; Chloridae; Gymnopus.
 ON NCBI_TaxID=160579;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hilu K.W., Alice L.A.;
 RT "A phylogeny of Chloridoideae (Poaceae) based on matK sequences."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS (BY SIMILARITY).
 CC -1- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED BY
 CC MITOCHONDRIAL INTRONS.
 DR EMBL: AF312333; AAK6024.1;
 DR InterPro: IPR000442; Intron_maturase2.
 DR Pfam: PF01348; Intron_maturase2.
 DR Pfam: PF01824; MatK_N; 1.
 KM Chloroplast; mRNA processing.
 SQ SEQUENCE 513 AA; 61432 MW; 56FAA4779579A0AF CRC64;

Query Match 5.2%; Score 99; DB 8; Length 513;

Best Local Similarity 19.3%; Pred. No. 3.3; 124; Indels 106; Gaps 18;
 Matches 70; Conservative 62; Mismatches 124; Indels 106; Gaps 18;

QY 22 QSLSLFQNSQSLOR---AYSTPYRYRIILOKENEKQALAHKICISILEFFKNLLFV 77
 DB 199 KSIFLPSKSKNKRSLFVLYNSVSEYSEFFLLFLRKOS-----SCLRLTSSGTF 246
 QY 78 HLISLSKQREGCSTDMAVSTPFENRLMY-----RLSSRSFL-----W 118
 DB 247 ERHIFSRKMEH-----FGVMYPGFEMKIWFEMDPLMHVYVQGYKYLAKGTLKKKW 301
 QY 119 KSYCPFFLDYLEAFGLSDPLDHOAVIKFELETHSSYYPVGFAPVQYSLDGRYF 178
 DB 302 KSLVNFSGYFSGFWTQGRIRLNLQNSGDFLGCHSSVPIITFLVNRQML-----ENFF 357
 QY 179 PLASVWRITDKNPSLTGDLIHDLGHVPMILHPSSEPFINNGRLFTK-----V 228
 DB 358 LIDTRMKFD-TTAPAT-----LIGSL-----SKAQTSGGHIKPIYADLSMDI 405
 QY 229 IEKVQAL-----PSKQRIQTLSNLIAIVRCFMTYVSGLIENHGRKAYGAVL 278
 DB 406 LDFGRIKCNLFHYHSSGSKQTYRLK-----YILR-----LSCAKTLARKHK-----STV 452
 QY 279 ISSPOLGAFIDNVRLPDLDOIIRLPNTSTPOETLFSIR--HFDELYELTSKLEWM 336
 DB 453 RFEWRLGSLVLEFPFT---BEDVFSILMF-----PKTHSPFGSHSERI-----WY 497
 QY 337 LD 338
 DB 498 LD 499

RESULT 43

054921 PRELIMINARY; PRT; 924 AA.

ID 054921
 AC 054921
 DT 01-JUN-1998 (TREMELREL. 06, Created)
 DT 01-JUN-1998 (TREMELREL. 06, Last sequence update)
 DT 01-DEC-2001 (TREMELREL. 19, Last annotation update)
 DE Rsees.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA MEDLINE=98070770; PubMed=9405631;
 RA Kee Y., Yoo J.S., Hazuka C.D., Peterson K.E., Hsu S.C., Scheller R.H.;
 RT "Subunit structure of the mammalian exocyst complex."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14438-14443(1997).
 DR EMBL: AF032666; AAC01578.1;
 DR InterPro: IPR002909; IPT_TIG.
 DR Pfam: PF01853; TIG; 1.
 SQ SEQUENCE 924 AA; 104031 MW; 1903C0593B113373 CRC64;

Query Match 5.2%; Score 99; DB 11; Length 924;

Best Local Similarity 19.9%; Pred. No. 6.7;
 Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;

QY 19 KLRQSLFQNSQSLORAYSTPYRYRIILOKENEKQALAHKICISILEFFKNLLFV 78
 DB 233 KVEGSMQKLEN--VNRASNTADTLFQEVIGRKD--ADSTRNALNVLRK-FLFVL 286
 QY 79 LLSLSKQREGCSTDMAVSTPFENRLMYRLSSR-----SLKSYCPFFLDYLEAF 133
 DB 287 PLNKRITQKG---DYDVIND-----YKASLFGKTEVGVFKY-----YAEV 329
 QY 134 GLSDPLDHOAVIKFELETHSSYYPVGFV-----APHQY-LSLIQ 174
 DB 330 AGIEDL--RELLKLU-LTPSTLHDQKRYIRYSDLHAPQDPAMOCIGAHKMTLKLMO 386
 QY 175 D-----RYPIASVWRITDKNPSLTGDLIHDLGHV----- 206
 DB 387 DCKGHNKSLKGNPGHSPW-----LDLDN-----DARPSVGHLSQFASLKGSSFGS 435
 QY 207 ---PWLH-----PSFSEPFINM--GRLETKYIEK--VQALPSK 238
 DB 436 GRDPTWRYKTPHRAVAVEKLTFLVLSQLPNFMKIMISYVNGSPSEFAESQGIERSKV 495
 QY 239 KOR-----IQTLQSNLIAIVR--CFMFTYVSGLIENHGRKAYGAVLS--SPQELG 286
 DB 496 RQRNDPKMKIQGVMSLVKYLIRGALLPFSIREG--DGRQ-YGVMEVQAEISQWLA 549
 QY 287 HAFIDNVV-----LPLEDOIIRLPNTSTPOETLFSIR-----HPEDEV 327
 DB 550 HV-IQTRLYESLTLAETIPNDMLQI-----QDLILDRHICIMWTLQHTAEIK 599
 QY 328 ELTSKLEWMDQGLSEIPLVNOEKYLSGEVL 360
 DB 600 RLAKEMWIDNDEGLTSLPCQFQSGIVASLSL 632

RESULT 44

08TPH6 PRELIMINARY; PRT; 1078 AA.

ID 08TPH6
 AC 08TPH6
 DT 01-JUN-2002 (TREMELREL. 21, Created)
 DT 01-JUN-2002 (TREMELREL. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)
 DE Hypothetical protein MA1936.
 GN MA1936.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=1193238;
 RA Galagan J.E., Nusbaum C., Roy A., Fitzhugh W., Calvo S., Engel R., Smirnov S., Arnoor D., Brown A.,
 Allen L., Nayler U., Stange-Thomann N., Dearellano K., Johnson R.,
 Linton L., McEwan P., McKernan K., Talamas J., Jitrell A., Ye W.,
 Zimmer A., Barber K.D., Cann I., Graham D.E., Graeme D.A., Guss A.M.,
 Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,

RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Unayam L.A., White O., White P.H., de Macario E.C.,
 RA Parry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Britchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.,
 RI "The Genome of Methanobrevibacterium smithii reveals extensive metabolic
 RI and physiological diversity.",
 RL Genome Res. 12:532-542(2002).
 RL EMBL; AB010876; AA05340.1; -.
 SW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1078 AA; 125949 MW; 9C99FDFB152B1351 CRC64;

Query Match 5.2%; Score 99; DB 17; Length 1078;
 Best Local Similarity 16.3%; Pred. No. 8;
 Matches 61; Conservative 79; Mismatches 148; Indels 86; Gaps 14;

QY 10 KYILKIALKRLQSLFQNSQS---LQRAYSTPSYVRIILQKENKEK-----QALA 60
 DB PHSLFESIPKFNKVMLLYKSDTEIEIYKIQNLSSAKIEIDNENYKALNLLDTEKK 195
 QY 61 RHKICISILEPFKNLLPVHLLSLSKNREGSCDMDVSTPFNRLWYLLSSRESLWKS 120
 DB 196 SDKCVDLQKHIDYIMNLK-LKLCENQ-----FENPENSFLKAQISV--- 236
 QY 121 YCRPFLLDYEARLLSDFLDQAVI-----KFFLETHFSYVPVSGFVAPHOYLQDR 176
 DB 237 -----TLEDFGI-----FLINVLKIEAKKIYEREIQLQDLNLYPENYEYLSLQSA 285
 QY 177 YFPIASVWRTLD-----KONFSLPDLIHLGHVPMLLHPSFSEF-----FINMGR 224
 DB 286 YNNGYVLSLGLIEBAKORYESQIHNLE-----IHPENEYQALTSAYYNLGNL 340
 QY 225 FTKV-----IEKQVALPSKQRIOTLQNLAIIVRCFWFTVESGLTEN 267
 DB 341 LSELGSDIENATNRYERALKVHTQIERYPEIKQYHSSVLNKEFKLIESFYCAENEI--N 398
 QY 268 HEGKAYGAVLITSSPOELGHAFI-----DNVRLVLELQIIRLPP-NTSPQTLPSI 320
 DB 399 NQTKWMPFGEVHMCEQYDLFIKSDSEDERKKMLFKIRSQIKFSFLDIEMQKHLSA 458
 QY 321 RHPDELVELTSKLE 334
 DB 459 EKCDIAKKIKKIK 472

RESULT 45
 QYITCO PRELIMINARY; PRT; 513 AA.

ID QYITCO
 AC QYITCO; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Probable intron maturase (Maturase K).
 GN MATK.
 OS Zeugites pittieri.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Centothecae; Zeugites.
 OC NCBI_TaxID=38726;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC HILU K.W., Alice L.A.;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Baeham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.,
 RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RA reveals hypervariable sequences.",
 RL Nature 403:665-668(2000).
 DR EMBL; ALI39079; CAB73666.1; -.
 DR InterPro; IPR001440; TPR.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 584 AA; 69439 MW; CFAD376B1562B1B CRC64;

Query Match 5.2%; Score 98.5; DB 16; Length 584;
 Best Local Similarity 20.2%; Pred. No. 4.2;
 Matches 81; Conservative 67; Mismatches 154; Indels 99; Gaps 19;

DR InterPro; IPR002866; Mark N.
 DR Pfam; PF01348; Iarxon maturase2; 1.
 DR Pfam; PF01824; Mark_N; 1.
 KW Chloroplast; rRNA processing.
 SQ SEQUENCE 513 AA; 60865 MW; BFDCA7C27343B94C CRC64;

Query Match 5.2%; Score 98.5; DB 8; Length 513;
 Best Local Similarity 22.0%; Pred. No. 3.6;
 Matches 88; Conservative 62; Mismatches 141; Indels 109; Gaps 25;

QY 1 VHCERTLPKYLKIALKRLQSLFQNSQS---RAYSTPSYVRIILQKENKEK-----IIL 49
 DB 147 LHLVSH-IEIPYPLHLEI-LVQILEYRIQDVPSLHLRFPFLNTYSNNWSLSSMKSILL 204
 QY 50 QKENKEQALARKHCISILEFFKNLLFVH-----LLSLSKN-----QREGCSTDM---AVVS 98
 DB 205 KKENKRLFRFLYNSVYSEYEF--LLFLRKQSSCLRTSSGTFLERIHCRKMEHFVVC 262
 QY 99 TPFENRLMY-----RLSSREFL-----WKSYPREFFLDYLEAFGLLSDP 139
 DB 263 PGFRKTIWFMDPLTHYVRYQKAILASKGTLKKKWSYLVN-FSQYFFSFGTQQR 321
 QY 140 LD-HQAVIKFEFELETHFSYVPVSGFVAPHOYLQDRYPIASVWRTLDKDNFSLTDL 198
 DB 322 IRLNLTNSCFDPLGYLSVPIITLVGNQML-----ENSLIDITRKMKFD-TTVPATP-- 374
 QY 199 IHDLLGHVPMLLHPSSEFFINMGRFTK-----VIEKQAL-----PSK 238
 DB 375 ---LIGSL-----SKAQCTGSGHPTSKPWTDLSDRDILDRFRCRNLFHYHSGSK 425
 QY 239 KQRIOTLQNLAIIVRCFWFTVESGLTENHEGRKAYGAVLITSSPOELGHAFIDNVRLPL 298
 DB 426 KQTLYRLK---YLRL--LSCARTLARKHK-----STVKTQMRLGVSFLEEFFT--- 469
 QY 299 ELQIIRLPNTSTPQTLFIRHFDLVELTSKLEWMLD 338
 DB 470 EEEQVFLMP---AKTTFSP-HGSQ-----SERIWYLD 499

RESULT 46
 QYPLZ8 PRELIMINARY; PRT; 584 AA.

ID QYPLZ8
 AC QYPLZ8; 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Hypothetical protein Cj1679.
 GN Cj1679.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OC NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Baeham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.,
 RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RA reveals hypervariable sequences.",
 RL Nature 403:665-668(2000).
 DR EMBL; ALI39079; CAB73666.1; -.
 DR InterPro; IPR001440; TPR.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 584 AA; 69439 MW; CFAD376B1562B1B CRC64;

Query Match 5.2%; Score 98.5; DB 16; Length 584;
 Best Local Similarity 20.2%; Pred. No. 4.2;
 Matches 81; Conservative 67; Mismatches 154; Indels 99; Gaps 19;

```

QY 10 PKYILKIALKLRQSLSPFONSOSLOPAYS-TPYSYRILLQXENKEKQALARKKICISIL 68
DB 29 PKFALFATLSSYKLSLFOKALYAAOELFSLNFTSPNGIMLAKSYTEN--LRDEALNML 86
QY 69 -----EPPKNLLFVHLLSLSKQRECGSTDMAVSTPPFNMLRYLLSSRFS 116
DB 87 QTLTRKDDLEDELKELAFIYKLSKNKLESEQIFKEL--LSKQMYNNL----- 133
QY 117 LMKSYCPREFL-DYLEAFG-----LSDPLDQAVIKFELFTHFSYVSGFVAPHQ 168
DB 134 LMKVAYEIFYGHDFTKALNHNHLCFHMOLDID-----KLQIAEQ 177
QY 169 YLSL--LQDRYFPIASVWRTLDKONFSLT---PDLIHDLGHVPLLHPSF----- 214
DB 178 NLNLVTKLEDR-----LHSTKENLTIKIEDFLTHQILPQKAVLLFKLFRISDSLFL 230
QY 215 -----SEFIMNGRLFTYVIE--KVQALPCKRQRIQTLQSNLIATVRCFPTVE 261
DB 231 OSLOEANOHAQFQWQON---YAKLEPNSNYQEAHYAKKLSLDSHATYQPDLAYLMLR 286
QY 262 SGLIEN-HEGRKAYGAVLISSPOELGHAFIDNVRLPLE--LDQIIR--LPRNTSTPQ 314
DB 287 MGVDDNFEQKKYKESRLFYAHNETFSTYHNESLKAENKFGVDAFKKKEVLPCEQGFQ 346
QY 315 ETLFSIRHDELVELTSLKEMMLDQGLLSIPLYNQKYL 355
DB 347 DTIMYARCLERKCKIASKVLPAPQAMYEMFK--NQIKFLN 385

```

RESULT 47

```

Q94DZ7 PRELIMINARY; PRT; 945 AA.
ID AC Q94DZ7.
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Putative zinc protease.
CN P0010B10.10.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 1, PAC
RT clone: P0010B10.10."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003224; BAB63566.1;
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 945 AA; 109356 MW; 0B46F635508477F6 CRC64;

```

Query Match 5.2%; Score 98.5; DB 10; Length 945;

Best Local Similarity 22.8%; Pred. No. 7.6; Mismatches 75; Indels 123; Gaps 19;

```

QY 27 FFONSQSLQRAYSTPIYSYRII-----LQENKEK----- 56
DB 442 FEGTDSVEWYCTAYSVENVTSMIQWIOKAPTEKLCIPKPNIFIKDSLKEAHKV 501
QY 57 -QALARKKICISILEFFKNLLF-----VHLL-----SLSKNREGSGTMAVSTPPFNRL 105
DB 502 KFPALIRKTPLSRLMWPMDLFTPKVHVIDHCPDLTSHSPE-----AVIST-----S 550
QY 106 LMYRLSSRFLMSKSYCPREFLDYLEAFG-----LSDPLDQAVIKFEL----- 151
DB 551 LFVDLLA-----DYLNAYGGYNDKMRKILDLAIMK-----ISNFVAYPNRMC 592

```

```

QY 152 ---ETHPSYVDSGVAPH---OYLSL-LQDRYFPIASVWRTLDKONFSLTLPDLIHLL 203
DB 593 ALKETAVKYOQNFKFSQPYQASNYLSLLEDDQWMPVEKLELRSK-----LEED--SLA 645
QY 204 GHVPWLLHPSFSEFFIMNGRL-----FTYVIEKVO-----ALPSKK--ORIQT 244
DB 646 KFIPIHLISKTFLEECYI-QGNIEPDNATISVQIEIDTIFNTPKSFVMSPOYLIRRVIT 704
QY 245 LQSNLIATVRCF 256
DB 705 LENEL-----KCY 712

```

RESULT 48

```

Q94D41 PRELIMINARY; PRT; 924 AA.
ID AC Q94D41.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 2410030124Rik protein.
CN Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakurai M., Straub J. P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Oikido T., Furuno M., Aono H., Baldarelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bortoldi M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita W., Gariboldi M.,
RA Garmuscin S., Hill D., Hotmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima U., Mazzerelli U., Mombaert P.,
RA Nordone P., Ring B., Ringwald W., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
DR EMBL; AK016532; BAB30290.1;
DR MGI; MGI:1913732; 2410030124Rik.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 1.
SQ SEQUENCE 924 AA; 103958 MW; 457BAD92EAA3040B CRC64;

```

Query Match 5.2%; Score 98; DB 11; Length 924;

Best Local Similarity 20.0%; Pred. No. 8.1; Mismatches 125; Indels 160; Gaps 26;

```

QY 19 KLRQSLSPFONSQSLQRAYSTPIYSYRIIQENKEKQALARKKICISILEFFKNLLFPH 78
DB 233 KVBGSMTQKLEN--VLRNSNADTLPOEVLGRQK--ADSTRNALNVQIRK-FLFNL 286
QY 79 LLSISKNREGSGTMAVSTPPFNRLMWRLSSRP-----SLMSYCPREFLDYLEAF 133
DB 287 PLNKRNQIGG--DYDVVIND-----YKASLSEKTEVOYFKKY-----YAEVE 329
QY 134 GLSDPLDQAVIKFELFTHPSYVDSGVF-----APHQY-LSLQ 174
DB 330 AGIEDL--RELILKKL-LTPSTLTDQKRYIRYLSDLHAGDDPAMQCIGAQHWTKLMQ 386
QY 175 D-----RYFPIASVWRTLDKONFSLTLPDLIHDLGHV----- 206

```


Db 387 DCKECHKMLKGGHPGPHSPM-----LDLDN-----DVRPSVLGHLSTQSLAKRGSSFQS 435
 QY 207 -----PWLH-----PSFSEFFINM--GRFTKVKIEKV-QALPSK-- 238
 Db 436 GRDDTWRYKTPHRVAFVEKLTUQLVQLFNWKLWISVYVGLSFSETAEKSGQSRKXV 495
 QY 239 -----KORIOTLOSNLIAIVCFWFTVESGLIENHGRKAYGAVLIS---SPOELGHA 288
 Db 496 RQRONDFFKMIQEVHSLVKLIRGALLPLS---LREGDGRQ-YGGEVQVQLSGQWLAVH 551
 QY 289 FIDNVRV-----LPLELDOIIRLPENTPTPOETLFSIR-----HPDELVEL 329
 Db 552 -IOTIRLTYESLTALEIPNDMLQII-----QDLIDIRICIMVTLOHTABEIKRL 601
 QY 330 TSKLEWMLDQGLLESIPLYNQEKYLSGFVL 360
 Db 602 ABKEDWVVDNEGLTSLPCQPEQSIHVSLSQSL 632

RESULT 49
 Q9GZG6 PRELIMINARY; PRT; 1887 AA.
 AC Q9GZG6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 214.2 kDa protein.
 GN F56A6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Murray J., Rohlfing T., O'Neal D., Wilson R.;
 RT "The sequence of C. elegans cosmid F56A6.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF067217; AAF99977.1; --
 DR HSSP; P08799; 1MND.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR000195; RA_domain.
 DR InterPro; IPR000198; RhogAP.
 DR Pfam; PF00130; DAG_PE-bind; 2.
 DR Pfam; PF00612; IQ; 4.
 DR Pfam; PF00663; myosin_head; 2.
 DR Pfam; PF00788; RA; 1.
 DR Pfam; PF06620; RhogAP; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00109; C1; 2.
 DR SMART; SM00015; IQ; 4.
 DR SMART; SM00242; MYSc; 1.
 DR SMART; SM00314; RA; 1.
 DR SMART; SM00324; RhogAP; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; UNKNOWN_2.

DR PROSITE; PS0081; DAG_PE_BIND_DOM_2; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 1887 AA; 214207 MW; 992DDE7F409D2D84 CRC64;
 Query Match 5.2%; Score 98; DB 5; Length 1887;
 Best Local Similarity 23.3%; Pred. No. 19;
 Matches 71; Conservative 34; Mismatches 92; Indels 108; Gaps 16;
 QY 9 DKFKILKALKRSLSLFQNSQSQRAYSTP-----YSYRIILLOKENKEQALA 60
 Db 217 NPKY-----ARLYFQS-----KRLGSLPHIFAIADVCHNMLRIKEN----- 254
 QY 61 RHKGISI-----LEFPKULFVHLISLKNQREGSCSTDMAVST-----PSFNRLWY 108
 Db 255 --QCVVIGSGSGKTESTNHLMSHLISLSQKSGTCTQLLSAGFVLEAFQNAVTLT 312
 QY 109 RLLSRFSLWKSYPFRF-LDYLEAFGLSLDLDHQAIVKFFLET-----H 154
 Db 313 NNNSSRFG-----KPIKINYREN-GWVSG-----ANVEIYLLEKRIIFQTKGERNYH 359
 QY 155 FSYYPVSG-----FVAPHQYLSLLQDRYFFPIASVMTLKDKNFSLTPDLIHLLG 204
 Db 360 VEYLLLEGADDEERKKYFLKPHDYKYLQNQNEFFALEGVN-----ERNEF-----DRLR 408
 QY 205 HVPMLLHPSFSEFFINMGRFTKVKIEKVALPSKKQRIQTQLQSNLIAIVRCFWFTVESGL 264
 Db 409 HA-----MSSVGFCAKTQQTIFGIISAV-----LLGNITVIKRHGYHSDSGY 452
 QY 265 IENHE 269
 Db 453 IENEE 457
 RESULT 50
 Q9NJQ3 PRELIMINARY; PRT; 239 AA.
 AC Q9NJQ3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Tryptophan hydroxylase (Fragment).
 OS Lymnaea stagnalis (Great pond snail).
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;
 OC Lymnaeidae; Lymnaea.
 OX NCBI_TaxID=6523;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21359676; PubMed=11466431;
 RA Koert C.E., Spencer G.E., van Minnen J., Li K.W., Geraerts W.P.,
 RA Syed N.I., Smit A.B., van Kesteren R.E.;
 RT "Functional implications of neurotransmitter expression during axonal
 RT regeneration; serotonin, but not peptides, auto-regulate axon growth
 RT of an identified central neuron.";
 RL J. Neurosci. 21:5597-5606(2001).
 DR EMBL; AF129815; AAF36488.1; --
 DR HSSP; P04176; 1PHZ.
 DR InterPro; IPR001273; Aaa_hydroxylase.
 DR Pfam; PF00351; biotin_H; 1.
 DR PRINTS; PR00372; FWHYDXYLASE.
 DR PROSITE; PS00367; BIOTFERIN_HYDROXYL; 1.
 DR NON_TER 1
 FT NON_TER 239
 SQ SEQUENCE 239 AA; 28001 MW; D721BD16C7D7B194 CRC64;
 Query Match 5.2%; Score 97.5; DB 5; Length 239;
 Best Local Similarity 21.7%; Pred. No. 1.7;
 Matches 34; Conservative 30; Mismatches 64; Indels 29; Gaps 5;
 QY 71 FKNLFLVHLISLKNQREGSCSTDMV-----VSTPFNRR-----NLWYRLSRFSLWKS 120
 Db 93 FKDTVY-----RQRKIITLAKYAHGTEIFVEYTEVEVETWGIIVFRELMLKLYPT 144
 QY 121 YCPFRFFLDYLEAFGLSLDLDH-----QAVIKFFLETHFSYYPVSGFVAPHQYLSL 172

Fri Jan 10 10:56:14 2003

us-09-438-185a-1047.rpt

Page 24

Db 145 HACR--EYLANIPFLVEHCGYREDNVPQLEDISRFLKERTGFTLRFVAGYLSRDPLAG 201
OY 173 LODRYFPPIASVWRTLDKDNFSLTPDLIHDLGHVPM 209
Db 202 LAFRVHCTQYIRHRDEPYTPPPCCHLMGMFLL 238

Search completed: January 9, 2003, 16:59:53
Job time : 90 secs

GenCore version 5.1.3
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 15:15:23 ; Search time 43 Seconds
(without alignments)
809,318 Million cell updates/sec

Title: US-09-438-185A-1047
Perfect score: 1889
Sequence: 1 VHYCERTLDPKYLKALKL.....ESIPLYNQEKYLSGPEVLQ 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	99.8	362	2 E72002	aromatic amino aci
2	1886	99.8	362	2 C85621	aromatic amino aci
3	256.5	13.6	289	2 D84113	phenylalanine-4-hy
4	235.5	12.5	262	2 F83535	phenylalanine-4-hy
5	231.5	12.3	444	1 S10489	tryptophan 5-mono
6	229.5	12.1	262	2 A53452	phenylalanine hydr
7	229.5	12.1	444	2 S51199	tryptophan 5-mono
8	229.5	12.1	491	2 A28582	tyrosine 3-mono
9	227.5	12.0	444	1 WHRTW	tryptophan 5-mono
10	226.5	12.0	447	2 A34582	tryptophan 5-mono
11	225.5	11.9	498	2 JN0068	tyrosine 3-mono
12	223.5	11.8	481	2 WHRTY	tryptophan 5-mono
13	223	11.8	498	1 WHRTY	tryptophan 5-mono
14	221.5	11.7	491	2 WHRTY	tyrosine 3-mono
15	220	11.6	528	1 WHRTY	tyrosine 3-mono
16	219	11.6	579	1 WHRTY	tyrosine 3-mono
17	214	11.3	453	1 WHRTY	phenylalanine 4-mo
18	213.5	11.3	453	1 WHRTY	phenylalanine 4-mo
19	212.5	11.2	453	2 JN0039	phenylalanine 4-mo
20	211.5	11.2	453	2 JN0039	phenylalanine 4-mo
21	207	11.0	433	2 JQ0766	phenylalanine 4-mo
22	206.5	10.9	575	2 JQ0766	phenylalanine 4-mo
23	197	10.4	453	1 A42271	hypothetical prote
24	194	10.3	452	2 JC4888	phenylalanine 4-mo
25	189.5	10.0	433	2 JC4888	phenylalanine 4-mo
26	178.5	9.4	404	2 T23494	phenylalanine 4-mo
27	175.5	9.3	457	2 T23494	phenylalanine 4-mo
28	109.5	5.8	1253	2 T45787	disease resistance
29	108.5	5.7	1501	1 B29813	174K ninaC protein

30	106.5	5.6	296	2 A40996	phenylalanine 4-mo
31	104.5	5.5	2331	2 S44054	genome polyprotein
32	102	5.4	563	2 T38766	probable transcrip
33	102	5.4	1050	2 C81624	exodeoxyribonuclea
34	100.5	5.3	1157	2 S38160	NUP133 protein - y
35	99	5.2	924	2 T09220	exocyst complex pr
36	98.5	5.2	584	2 D81265	hypothetical prote
37	98.5	5.2	1024	2 C64208	hypothetical prote
38	98	5.2	1846	2 T33079	hypothetical prote
39	97	5.1	492	1 S88856	cytochrome P450 2L
40	97	5.1	1050	2 G86582	exodeoxyribonuclea
41	96.5	5.1	512	2 F86520	glucose-6-p dehydro
42	96.5	5.1	512	2 C72103	glucose-6-p dehydro
43	96	5.1	529	2 S62468	probable membrane
44	96	5.1	1050	2 H72041	exodeoxyribonuclea
45	96	5.1	2059	2 T41933	large tegument pro
46	95.5	5.1	536	2 E70066	hypothetical prote
47	95	5.0	970	2 B84846	probable zinc prot
48	95	5.0	1427	2 S74293	SR88 protein - yea
49	95	5.0	1475	2 A60026	cell communication
50	94.5	5.0	599	2 C81321	probable cation-tr
51	94	5.0	293	2 F71949	hypothetical prote
52	94	5.0	317	2 F71139	hypothetical prote
53	94	5.0	618	2 E72285	ABC transporter, p
54	93	4.9	307	2 D86838	hypothetical prote
55	93	4.9	645	2 T25824	hypothetical prote
56	93	4.9	714	2 C90100	developmental prote
57	93	4.9	1022	2 T17406	hypothetical prote
58	93	4.9	1391	2 T20642	hypothetical prote
59	93	4.9	1397	2 B87998	heterocyst maturat
60	92.5	4.9	466	2 B64207	hypothetical prote
61	92.5	4.9	735	2 T39464	hypothetical prote
62	92.5	4.9	1084	2 C92931	hypothetical prote
63	92	4.9	447	2 S28086	gene A protein - y
64	92	4.9	523	2 S50479	26S proteasome reg
65	92	4.9	544	2 S58532	matK protein (trnk
66	92	4.9	520	2 JCS567	lamina B receptor -
67	92	4.9	4563	1 LPHUB	apolipoprotein B-1
68	91.5	4.8	293	2 B64339	apelinoprotein B-1
69	91.5	4.8	670	2 G90091	spermidine synthas
70	91.5	4.8	907	2 A86460	hypothetical prote
71	91	4.8	935	1 FXR5R	93.9K hypothetical
72	91	4.8	1397	2 T10466	minor inner core p
73	91	4.8	2183	1 ZLNZMV	DNA topoisomerase
74	91	4.8	2183	1 G48556	genome polyprotein
75	90.5	4.8	314	2 A84960	probable signal pe
76	90.5	4.8	616	2 S64746	ORC3 protein - yea
77	90.5	4.8	1102	2 S55100	hypothetical prote
78	90.5	4.8	2330	1 RRIWMV	genome polyprotein
79	90	4.8	863	2 S37088	phycobilisome anch
80	90	4.8	892	2 T15760	hypothetical prote
81	90	4.8	2692	2 T23768	hypothetical prote
82	89.5	4.7	519	2 T28260	hypothetical prote
83	89.5	4.7	553	2 D98026	hypothetical prote
84	89.5	4.7	2109	2 T31352	hypothetical prote
85	89.5	4.7	2136	2 A05037	hypothetical prote
86	89.5	4.7	2493	2 T40540	hypothetical prote
87	89.5	4.7	2812	2 T43271	phosphotidylinosi
88	89	4.7	355	2 G90513	hypothetical prote
89	89	4.7	374	2 S50372	hypothetical prote
90	89	4.7	1037	2 T26200	hypothetical prote
91	89	4.7	1397	1 G85010	hypothetical prote
92	89	4.7	1397	2 A51035	sensor protein avg
93	89	4.7	1197	2 C85879	probable sensor fo
94	88.5	4.7	519	2 T34509	guanine nucleotide
95	88.5	4.7	892	2 F64703	DNA polymerase I -
96	88.5	4.7	1967	2 D64604	hypothetical prote
97	88.5	4.7	439	2 S64510	hypothetical prote
98	88	4.7	743	2 D82883	DNA helicase II UU
99	88	4.7	873	2 S46584	probable membrane
100	88	4.7	873	2 S46584	polymerase - Berne
101	88	4.7	2291	2 S11238	apolipoprotein B -
102	88	4.7	2629	2 I46569	apolipoprotein B -

103	87.5	4.6	334	2	E71680	cytochrome C oxid
104	87.5	4.6	454	2	AH0469	probable 4-alpha-L
105	87.5	4.6	542	2	J00201	hypothetical 65K p
106	87.5	4.6	694	2	D81280	polysphosphate kina
107	87.5	4.6	1122	2	A97814	transcription-repa
108	87.5	4.6	1237	2	AC1583	internalin protein
109	87.5	4.6	1817	2	H72611	probable secreted
110	87	4.6	239	2	F64558	conserved hypotet
111	87	4.6	324	1	A49853	choylglycine hyd
112	87	4.6	741	2	H90507	atp-dependent heli
113	86.5	4.6	128	2	AH2322	hypothetical prote
114	86.5	4.6	533	2	T27589	hypothetical prote
115	86.5	4.6	710	2	B89252	protein ZC455.4 (l
116	86.5	4.6	990	2	S23416	lanthibiotic epide
117	86.5	4.6	1202	2	A48773	ribonuclease P (EC
118	86	4.6	339	2	T20860	hypothetical prote
119	86	4.6	422	2	A72309	conserved hypotet
120	86	4.6	512	2	D98074	conserved hypotet
121	86	4.6	565	2	D72222	conserved hypotet
122	86	4.6	1405	1	DJ2BA	DNA-directed DNA p
123	86	4.6	1822	2	S44849	K12H4.8 protein -
124	85.5	4.5	321	2	D90154	transposase ISC143
125	85.5	4.5	816	2	T25788	hypothetical prote
126	85.5	4.5	1489	2	T38842	probable RAS GTPas
127	85.5	4.5	1613	2	D90129	hypothetical prote
128	85.5	4.5	1808	2	AB1847	serine/threonine k
129	85.5	4.5	2039	2	S64540	probable calcium c
130	85	4.5	238	1	B64313	probable 3-isoprop
131	85	4.5	432	1	T32210	hypothetical prote
132	85	4.5	462	1	YS8YTM	threonine-RNA lig
133	85	4.5	471	2	F71618	adenylomycinase I
134	85	4.5	493	2	H95209	transcription acti
135	85	4.5	549	2	B64540	conserved hypotet
136	85	4.5	567	2	B69166	hypothetical prote
137	85	4.5	717	2	S78177	hypothetical prote
138	85	4.5	938	2	G70472	hypothetical prote
139	84.5	4.5	358	2	T22823	hypothetical prote
140	84.5	4.5	456	2	G71152	hypothetical prote
141	84.5	4.5	531	2	B84442	hypothetical prote
142	84.5	4.5	553	2	E95160	hypothetical prote
143	84.5	4.5	585	2	S48951	hypothetical prote
144	84.5	4.5	719	2	C88216	protein B0495.2 (l
145	84.5	4.5	820	1	T27909	hypothetical prote
146	84.5	4.5	1070	2	G84982	exodeoxyribonuclea
147	84.5	4.5	1107	2	S52517	myosin I heavy cha
148	84.5	4.5	1748	1	JQ1555	genome polyprotein
149	84.5	4.5	3068	1	A44062	genome polyprotein
150	84	4.4	240	2	F70409	conserved hypotet

ALIGNMENTS

RESULT 1
 E72002
 aromatic amino acid hydroxylase - Chlamydomonas pneumoniae (strain CML029)
 C/Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C/Accession: E72002
 R/Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A/Reference number: A72000; MUID:92060606; PMID:10192388
 A/Accession: E72002
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-362 <ARN>
 A/Cross-references: GB:AE001685; GB:AE001363; NID:g4377378; PIDN:AAD19183.1; PID:g437738
 A/Experimental source: strain CML029
 C/Genetics:
 A/Gene: CP1046
 Query Match 99.8%; Score 1886; DB 2; Length 362;

Best Local Similarity 99.7%; Pred. No. 2e-144; Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	VHVCERTLPKYLILKALKQSLSPFQNSQSLQRYSTPYSYRIILQKNEKQOLA	60						
Db	1	MHVCERTLPKYLILKALKQSLSPFQNSQSLQRYSTPYSYRIILQKNEKQOLA	60						
Qy	61	RHKCISLLEFPKLLFVHLISLKNQRCGCTDAVAVSTFFRNIMYRLSSRSLWKS	120						
Db	61	RHKCISLLEFPKLLFVHLISLKNQRCGCTDAVAVSTFFRNIMYRLSSRSLWKS	120						
Qy	121	YCRPFLDYLEAFGLSDPLDQAVIKFELETHFSYYPVSGVAPHOYLSTLDQRYPI	180						
Db	121	YCRPFLDYLEAFGLSDPLDQAVIKFELETHFSYYPVSGVAPHOYLSTLDQRYPI	180						
Qy	181	ASVWRFLDQNSLTPDILHDLGHVPLHPSPSEFFIMGRLEFKVIEKVALPSKKQ	240						
Db	181	ASVWRFLDQNSLTPDILHDLGHVPLHPSPSEFFIMGRLEFKVIEKVALPSKKQ	240						
Qy	241	RIQTLOSNLAIYRCFWFTVESGLIENHGRKAGAVLISPOLGHAFIDNVRVPLEL	300						
Db	241	RIQTLOSNLAIYRCFWFTVESGLIENHGRKAGAVLISPOLGHAFIDNVRVPLEL	300						
Qy	301	DOIRLPMTSTPOETLSIRHPELVELTSKLEWMDQGLSIPLYNQEKYLSGFVYL	360						
Db	301	DOIRLPMTSTPOETLSIRHPELVELTSKLEWMDQGLSIPLYNQEKYLSGFVYL	360						
Qy	361	CO 362							
Db	361	CO 362							
RESULT 2									
aromatic amino acid hydroxylase (imported) - Chlamydomonas pneumoniae (strain J138)									
C/Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae									
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001									
C/Accession: C86621									
R/Shiira, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.									
Nucleic Acids Res. 28, 2311-2314, 2000									
A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.									
A/Reference number: A86491; MUID:20330349; PMID:10871362									
A/Accession: C86621									
A/Status: preliminary									
A/Molecule type: DNA									
A/Residues: 1-362 <STO>									
A/Cross-references: GB:BA000008; NID:98979419; PIDN:BA09253.1; GSPDB:GN00142									
A/Experimental source: strain J138									
C/Genetics:									
A/Gene: CP1046									
Query Match 99.8%; Score 1886; DB 2; Length 362;									
Best Local Similarity 99.7%; Pred. No. 2e-144; Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	VHVCERTLPKYLILKALKQSLSPFQNSQSLQRYSTPYSYRIILQKNEKQOLA	60						
Db	1	MHVCERTLPKYLILKALKQSLSPFQNSQSLQRYSTPYSYRIILQKNEKQOLA	60						
Qy	61	RHKCISLLEFPKLLFVHLISLKNQRCGCTDAVAVSTFFRNIMYRLSSRSLWKS	120						
Db	61	RHKCISLLEFPKLLFVHLISLKNQRCGCTDAVAVSTFFRNIMYRLSSRSLWKS	120						
Qy	121	YCRPFLDYLEAFGLSDPLDQAVIKFELETHFSYYPVSGVAPHOYLSTLDQRYPI	180						
Db	121	YCRPFLDYLEAFGLSDPLDQAVIKFELETHFSYYPVSGVAPHOYLSTLDQRYPI	180						
Qy	181	ASVWRFLDQNSLTPDILHDLGHVPLHPSPSEFFIMGRLEFKVIEKVALPSKKQ	240						
Db	181	ASVWRFLDQNSLTPDILHDLGHVPLHPSPSEFFIMGRLEFKVIEKVALPSKKQ	240						
Qy	241	RIQTLOSNLAIYRCFWFTVESGLIENHGRKAGAVLISPOLGHAFIDNVRVPLEL	300						
Db	241	RIQTLOSNLAIYRCFWFTVESGLIENHGRKAGAVLISPOLGHAFIDNVRVPLEL	300						
Qy	301	DOIRLPMTSTPOETLSIRHPELVELTSKLEWMDQGLSIPLYNQEKYLSGFVYL	360						
Db	301	DOIRLPMTSTPOETLSIRHPELVELTSKLEWMDQGLSIPLYNQEKYLSGFVYL	360						
Qy	361	CO 362							
Db	361	CO 362							

DB 241 RIOTLQNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISPPQELGHAFIDNVRVLPLEL 300
QY 301 DOIIRLPNTSTPQETLFSIRHFDLVELTSLKLEMLDQGLLESIPLYNQEKYLSGFEVL 360
DB 301 DOIIRLPNTSTPQETLFSIRHFDLVELTSLKLEMLDQGLLESIPLYNQEKYLSGFEVL 360
QY 361 CQ 362
DB 361 CQ 362

RESULT 3
phenylalanine-4-hydroxylase VCA0828 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: D82413
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chaudhuri, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, J.; R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: D82413
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-289 <HEI>
A:Cross-references: GB:AE004410; GB:AE003853; NID:99658244; PIDN:AAF96726.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0828
A:Map position: 2

Query Match 13.6%; Score 256.5; DB 2; Length 289;
Best Local Similarity 25.6%; Pred. No. 3.2e-13;
Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;

QY 106 LWYLLSRFLSKSCYCPREFDYLEAFOLLSDFLDHOAVT-KFELETHFSYVSGPV 164
DB 49 VHELTIRQEEVTRACQAYLDGLMLPDLRPLQPEINRVLRQETGWQVEVPALI 108
QY 165 APHQVSLQDQRYPIASVNRDLDKONFSLTDLHLDLGHVPLHPSFSEFFINMRL 224
DB 109 SDRFALLADKFPVATFRRREEDFYLDQEDPFHEVYGHCAMLTHPDFAAFTHVYGO 168
QY 225 FTKVIEKQALPSKKQRIOTLQNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISPPQ 284
DB 169 GAKATPKERSY-----LARYWFTVEFGLVQEQCKTIYGGGILSSPGE 212
QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSLK-----LEWM-LD 338
DB 213 TLVASESTIPKREFPDIMQVLRTPYRIDIMQPIYVYVLPDLSQLYOLSORDVMALVWQMQ 272
QY 339 QGLLESIPLYNQEK 352
DB 273 DGLLP--PLFQPK 284

RESULT 4
phenylalanine-4-hydroxylase PA0872 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83535
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83535
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-262 <STO>
A:Cross-references: GB:AE004522; GB:AE004091; NID:99946768; PIDN:AAG04261.1; GSPDB:GN001;
A:Experimental source: strain PA01
C:Genetics:
A:Gene: phbA; PA0872

Query Match 12.5%; Score 235.5; DB 2; Length 262;
Best Local Similarity 24.5%; Pred. No. 1.4e-11;
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;

QY 93 DMVVSTPFNNRLWYLLSRFLSKSCYCPREFDYLEAFOLLSDFLDHOAVIKFPELE 152
DB 13 DNGFIHPETEHOQVMTLITRQKLVIEGRACQEVLDGIEQLG-----LPHERIPQLDEIN 67
QY 153 -----THFSYVSGFVAPHOYLSILODRYEPITASVNRDLDKONFSLTDLHLDLGHV 206
DB 68 RVLAQTGWRVARVPALIPFOFFELLASQFFVATIRTFPELDYLOEPDIFHEIFGHC 127
QY 207 PHLHPSFSEFFINMRLTQVIEKQALPSKKQRIOTLQNLIAIVRCFWFTVESGLIE 266
DB 128 PLLNPFPAETHTYKGLKA-----SKEER-----VFLARYWMTIEFGLVE 171
QY 267 NHEGRKAYGAVLISPPQELGHAFID---NVRVLPLELDQIIRLPNTSTPQETLFSIRHF 323
DB 172 TDQGRYGGGILSSPKETVYSLSDPLHOAFNPLE---AMRTPYRIDILQPLVFLPD 228
QY 324 DELVELTSLK 332
DB 229 KRLFLAQE 237

RESULT 5
S10489
tryptophan 5-monoxygenase (EC 1.14.16.4) - human
N:Alternate names: tryptophan 5-hydroxylase
C:Species: Homo sapiens (nan)
C>Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 03-Mar-2000
C:Accession: S10489; S51559
R:Soularand, S.; Darmon, M.C.; Ganem, Y.; Launay, J.M.; Mallet, J.
Nucleic Acids Res. 18, 4257, 1990
A:Title: Complete coding sequence of human tryptophan hydroxylase.
A:Reference number: S10489; MUID:90332431; PMID:2377472
A:Accession: S10489
A:Molecule type: mRNA
A:Residues: 1-444 <BOU>
A:Cross-references: EMBL:X52836; NID:937954; PIDN:CAA37018.1; PID:G37955
R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.
Arch. Biochem. Biophys. 315, 445-453, 1994
A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in
A:Reference number: S51199; MUID:95077422; PMID:7986090
A:Accession: S51559
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-18, 'T', 20-67, 'T', 69-89, 'TP', 92-96, 'M', 98-99, 'E', 101-103, 'S', 105-150, 'S', 15;
35, 'G', 437-444 <TIP>
A:Cross-references: GB:L29306; NID:9531192; PIDN:AAA67050.1; PID:9531193
C:Genetics:
A:Gene: GDB:TPH; TPH
A:Cross-references: GDB:L20732; OMIM:191060
A:Map position: lp15.1-lp14.3
C:Function:
A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahyd;
A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism
C:Superfamily: phenylalanine 4-monoxygenase
C:Keywords: bioprotein; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxid;
F58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict;
F260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat;
F272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.3%; Score 231.5; DB 1; Length 444;
Best Local Similarity 27.8%; Pred. No. 5.8e-11;
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;

DB 343 QDIG-----LASLGATDEIEKLAFL-----YWFVEFLGCRQNGIVKAYGAGL 386
QY 279 ISSQELGHAFIDNVRLVPLELQIIRLPNTSTPQETLFSIRHFDE 325
DB 387 LSSYGELIHSLSDEPEVRDPDAAAQVPCDQPYQPVYFVSESFS 433

RESULT 9

WRTW

N:Altermate names: tryptophan 5-hydroxylase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 03-Mar-2000
C:Accession: J00034; A60034; A24367
R:Dammon, M.C.; Guilbert, B.; Leviel, V.; Ehret, M.; Maitre, M.; Mallet, J.
J. Neurochem. 51, 312-316, 1988
A:Title: Sequence of two mRNAs encoding active rat tryptophan hydroxylase.
A:Reference number: J00034; MUID:88244702; PMID:3379411
A:Accession: J00034
A:Molecule type: mRNA
A:Residues: 1-444 <DAR>
A:CROSS-references: GB:X53501; NID:g57760; PIDN:CAA37579.1; PID:g57761
A:Experimental source: pineal gland
R:Kim, K.S.; Wessel, T.C.; Stone, D.M.; Carver, C.H.; Joh, T.H.; Park, D.H.
Brain Res. Mol. Brain Res. 9, 277-283, 1991
A:Title: Molecular cloning and characterization of cDNA encoding tryptophan hydroxylase
A:Reference number: A60034; MUID:91245924; PMID:1645430
A:Accession: A60034
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-444 <KIM>
A:Experimental source: dorsal raphe nucleus
R:Damon, M.C.; Grime, B.; Cash, C.D.; Maitre, M.; Mallet, J.
FEBS Lett. 206, 43-46, 1986
A:Title: Isolation of a rat pineal gland cDNA clone homologous to tyrosine and phenylala
A:Reference number: A24367; MUID:87005247; PMID:2875901
A:Accession: A24367
A:Molecule type: mRNA
A:Residues: 167-261 <DA2>
A:CROSS-references: GB:M28000; NID:g207432; PIDN:AAA42262.1; PID:g207433
C:Comment: This enzyme has different physical properties in pineal gland and in dorsal h
C:Function:
A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahyd
A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxid
F:58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict
F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #sta
F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 1.2e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;
QY 107 WYRLSSRFLSKSYCRFFLDYLEAFGLLSDFLDH-----QAVIKFFELTHFSY 158
DB 174 WGTIFRELKLYPTHACR---EYLRNLPLLSKYCYREDNPQLQEDVSNFLKERTGFSIR 230
QY 159 PVSGFVAPHOYVLLQDRYFPIASVVRTLDKDNFSLTPDLIHLGHVPMWLLHPSFSFF 218
DB 231 PVAGYLSPRDFLSGLAFRVFCHTQYVVRHSSDPLYTPEPTCHELLGHVPLLAEPSPAQS 290
QY 219 INMGRLTKVIEKQVQALPSKKORIOTLQSNLIAIVRCFWFTVSGLIENHEGRKAYGAVL 278
DB 291 QBIG-----LASLGASEETVOKLAT-----CYFFTFVEFLGCKQDGLRVFGAGL 334
QY 279 ISSQELGHAFIDNVRLVPLELQIIRLPNTSTPQETLFSIRHFDELVE 328
DB 335 LSSISELRHLSGHAKVPDPKVPACKOECLITSFDQVYFVSESFEDEAKE 384

RESULT 10

A34582
tryptophan 5-monooxygenase (EC 1.14.16.4) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jun-1990 #sequence_revision 09-Oct-1992 #text_change 31-Mar-2000
C:Accession: A34582
R:Stoll, J.; Kozak, C.A.; Goldman, D.
Genomics 7, 88-96, 1990
A:Title: Characterization and chromosomal mapping of a cDNA encoding tryptophan hydroxylase
A:Reference number: A34582; MUID:90243261; PMID:2110547
A:Accession: A34582
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-447 <STO>
A:CROSS-references: GB:J04758; NID:g202113; PIDN:AAA63401.1; PID:g202114
A:Note: the authors translated the codon AAC for residue 405 as Gln
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; iron; metalloprotein; oxidoreductase; phosphoprotein
F:275,280,320/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 226.5; DB 2; Length 447;
Best Local Similarity 27.4%; Pred. No. 1.5e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;
QY 107 WYRLSSRFLSKSYCRFFLDYLEAFGLLSDFLDH-----QAVIKFFELTHFSY 158
DB 177 WGTIFRELKLYPTHACR---EYLRNLPLLSKYCYREDNPQLQEDVSNFLKERTGFSIR 233
QY 159 PVSGFVAPHOYVLLQDRYFPIASVVRTLDKDNFSLTPDLIHLGHVPMWLLHPSFSFF 218
DB 234 PVAGYLSPRDFLSGLAFRVFCHTQYVVRHSSDPLYTPEPTCHELLGHVPLLAEPSPAQS 293
QY 219 INMGRLTKVIEKQVQALPSKKORIOTLQSNLIAIVRCFWFTVSGLIENHEGRKAYGAVL 278
DB 294 QEIG-----LASLGASEETVOKLAT-----CYFFTFVEFLGCKQDGLRVFGAGL 337
QY 279 ISSQELGHAFIDNVRLVPLELQIIRLPNTSTPQETLFSIRHFDELVE 328
DB 338 LSSISELRHLSGHAKVPDPKVPACKOECLITSFDQVYFVSESFEDEAKE 387

RESULT 11

JN0068
tyrosine 3-monooxygenase (EC 1.14.16.2) - mouse
N:Altermate names: tyrosine hydroxylase
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Mar-2000
C:Accession: JN0068; S21322
R:Ichikawa, S.; Sasaoka, T.; Nagatsu, T.
Biochem. Biophys. Res. Commun. 176, 1610-1616, 1991
A:Title: Primary structure of mouse tyrosine hydroxylase deduced from its cDNA.
A:Reference number: JN0068; MUID:91248263; PMID:1674869
A:Accession: JN0068
A:Molecule type: mRNA
A:Residues: 1-498 <ICH>
A:CROSS-references: GB:M69200; NID:g201997; PIDN:AAA40434.1; PID:g201998
R:Morgan, W.W.; Bermudez, J.; Sharp, Z.D.
Submitted to the EMBL Data Library, June 1990
A:Description: DC-12 Nuclear Extracts Produce Tissue-Specific Protection of Several Sequ
A:Reference number: S21322
A:Accession: S21322
A:Molecule type: DNA
A:Residues: 1-30 <MOR>
A:CROSS-references: EMBL:X53503; NID:g55055; PIDN:CAA37580.1; PID:g55056
C:Comment: This enzyme, which requires ferrous iron, catalyzes the hydroxylation of tyros
role in the physiology of adrenergic neurones.
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bipterin; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein
F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.9%; Score 225.5; DB 2; Length 498;
Best Local Similarity 27.5%; Pred. No. 2.1e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;

Qy 107 WYRLSSRFSLMKSCYCRFPDLYLEAFGLSDPLDH-----QAVIKPELETHNSY 158
 Db 233 MEKVATYTKGLKATPHACK--EHLAEQLEKICGIRREDSIQOLEVSHFLKRGQLR 289
 Qy 159 PVSGFVAAPHQYLSLDQRYFPFASVMTLDKNFSLTPDLIHLLGHVPMLLHPSSEFP 218
 Db 290 PVAGLLSARDPLASLAPFVQCQOYIRHASSPMHSEPPCCHELLGHVPMADRTFAQFS 349
 Qy 219 IMNGRLFTKVIKVKQALPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYGAVL 278
 Db 350 QDIG-----LASLGASDEELIKLST-----VWFVTEFGICKQNGELKAYGGL 393
 Qy 279 ISSPOLGHAFTDNVRLPLELDQIRLPNTSTPOETLFSIRHF---DELVELTSKLE 334
 Db 394 ISSYGEHLHSLSEBEVRAFPDPAVQYQDQYQPVYFVSESFSDAKDKLRNYSRIQ 453

RESULT 12

151567
 tryptophan 5-monooxygenase (EC 1.14.16.4) - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 31-Mar-2000
 C:Accession: 151567
 R:Green, C.B.; Beehars, J.C.
 J. Neurochem. 62, 2420-2428, 1994
 A:Title: Tryptophan hydroxylase expression is regulated by a circadian clock in Xenopus
 A:Reference number: 151567; MUID:94246419; PMID:8189245
 A:Accession: 151567
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-481 <GRE>
 A:Cross-references: GB:L20679; NID:9450644; PIDN:AAA21306.1; PID:9450645
 C:Superfamily: phenylalanine 4-monooxygenase
 C:Keywords: bioprotein, iron, metalloprotein, oxidoreductase
 F:309,314,354/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223.5; DB 2; Length 481;
 Best Local Similarity 29.5%; Pred. No. 2.9e-10;
 Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;
 Qy 107 WYRLSSRFSLMKSCYCRFPDLYLEAFGLSDPLDH-----QAVIKPELETHNSY 158
 Db 211 WCTVRELANKYPTHACK--EYKULPLSKKCGIRENITQLEVDVSFLKERTGTLIR 267
 Qy 159 PVSGFVAAPHQYLSLDQRYFPFASVMTLDKNFSLTPDLIHLLGHVPMLLHPSSEFP 217
 Db 266 PVAGLLSARDPLASLAPFVQCQOYIRHASSPMHSEPPCCHELLGHVPMADRTFAQFS 326
 Qy 219 IMNGRLFTKVIKVKQALPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYGAVL 277
 Db 327 SOEIG-----LASLGASDEAVQKLTAT-----CYFVTEFGICKQNGELKAYG 370
 Qy 279 ISSPOLGHAFTDNVRLPLELDQIRLPNTSTPOETLFSIRHFDELVELTSKLE 334
 Db 394 ISSYGEHLHSLSEBEVRAFPDPAVQYQDQYQPVYFVSESFSDAKDKLRNYSRIQ 453

RESULT 13

WHRTY
 tyrosine 3-monooxygenase (EC 1.14.16.2) - rat
 N:Alternate names: tyrosine 3-hydroxylase
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 03-Mar-2000
 C:Accession: A00510; A44714; S03026; I58264
 R:Germa, B.; Lamouroux, A.; Bianot, F.; Faucon Biguet, N.; Mallet, J.
 Proc. Natl. Acad. Sci. U.S.A. 82, 617-621, 1985
 A:Title: Complete coding sequence of rat tyrosine hydroxylase mRNA.
 A:Reference number: A00510; MUID:85113249; PMID:2857492
 A:Accession: A00510
 A:Molecule type: mRNA
 A:Residues: 1-498 <GRI>
 A:Cross-references: GB:M10244; NID:g207408; PIDN:AAA42257.1; PID:g207409

R:Campbell, D.G.; Hardie, D.G.; Vulliam, P.R.
 J. Biol. Chem. 261, 10489-10492, 1986
 A:Title: Identification of four phosphorylation sites in the N-terminal region of ty
 A:Reference number: A44714; MUID:86278113; PMID:2874140
 A:Accession: A44714
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 2-12,16-24,38-47,151-157 <CAN>
 R:Bonnefoy, E.; Ferrara, P.; Rohrer, H.; Gros, F.; Thibault, J.
 Eur. J. Biochem. 174, 685-690, 1988
 A:Title: Role of the N-terminus of rat pheochromocytoma tyrosine hydroxylase in the
 A:Reference number: S03026; MUID:88271342; PMID:2899026
 A:Accession: S03026
 A:Molecule type: protein
 A:Residues: 2-26 <BON>
 R:Harrington, C.A.; Lewis, E.J.; Krzemien, D.; Chikaraishi, D.M.
 Nucleic Acids Res. 15, 2363-2384, 1987
 A:Title: Identification and cell type specificity of the tyrosine hydroxylase gene p
 A:Reference number: I58264; MUID:87174758; PMID:2882469
 A:Accession: I58264
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-30 <RBS>
 A:Cross-references: EMBL:X04914; NID:957355; PIDN:CAA28584.1; PID:957356
 C:Function:
 A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylal
 A:Pathway: catecholamine biosynthesis
 A:Note: this is the rate-limiting step in catecholamine biosynthesis
 C:Superfamily: phenylalanine 4-monooxygenase
 C:Keywords: bioprotein; catecholamine biosynthesis; iron; metalloprotein; monooxyg
 F:8/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status experi
 F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #stat
 F:40,133/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #statu
 F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223; DB 1; Length 498;
 Best Local Similarity 26.7%; Pred. No. 3.3e-10;
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;
 Qy 112 SSRFSLMKS-----XCRFPDLYLEAFGLSDPLDH-----QAVIKPELETHNF 155
 Db 227 AEARATKVEYVYTKQVATYACREHLGFLERKGYGVREDSIPQLEVDVSFLKERTCF 286
 Qy 156 SYTVSGFVAAPHQYLSLDQRYFPFASVMTLDKNFSLTPDLIHLLGHVPMLLHPSFS 215
 Db 287 QLRVAGLLSARDPLASLAPFVQCQOYIRHASSPMHSEPPCCHELLGHVPMADRTFA 346
 Qy 219 IMNGRLFTKVIKVKQALPSKKQRIQTLQSNLAIIVRCFMTVESGLIENHGRKAYG 275
 Db 347 QFQDQIG-----LASLGASDEELIKLST-----VWFVTEFGICKQNGELKAYG 390
 Qy 276 AVLISSPOLGHAFTDNVRLPLELDQIRLPNTSTPOETLFSIRHF---DELVELTS 331
 Db 391 AGLISSYGEHLHSLSEBEVRAFPDPAVQYQDQYQPVYFVSESFSDAKDKLRNYS 450
 Qy 332 KLE 334
 Db 451 RIQ 453

RESULT 14

145983
 tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
 C:Accession: I45983
 R:DeWello, S.R.; Weisberg, E.P.; Stachowiak, M.K.; Turzai, L.M.; Gloto, A.E.; Kaplan,
 J. Neurosci. Res. 19, 440-449, 1988
 A:Title: Isolation and nucleotide sequence of a cDNA clone encoding bovine adrenal ty
 A:Reference number: I45983; MUID:88259287; PMID:2898537
 A:Accession: I45983
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

C:Function:
 A:Description: catalyzes the 4'-hydroxylation of phenylalanine to tyrosine by tetrahydrot
 A:Pathway: tyrosine biosynthesis; phenylalanine catabolism
 C:Superfamily: phenylalanine 4-monooxygenase
 C:Keywords: bioprotein; homodimer; iron; metalloprotein; monooxygenase; oxidoreductase; p
 F:16/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status p
 F:23/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict
 F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.3%; Score 214; DB 1; Length 453;
 Best Local Similarity 28.4%; Pred. No. 1.5e-09;
 Matches 61; Conservative 37; Mismatches 89; Indels 28; Gaps 4;

QY 115 FSLMKSYPFRFLDYLEAFGLSDFLDH-QAVIKFELETHFSYYPVSGFVAPHQYLSLL 173
 DB 210 FLELEKYC-----GFREDNPQLEDVVSQFLQCTGFRLRPVAGLLSSRDLGL 258

QY 174 QDRYPTASVMRTLDKNFSLDPLDHLHGVPLWLLHPSSEFFINMGRLETKVIEKVQ 233
 DB 259 APRVPHCTQYIRHSGKPMYTPEDPICHLLGHVPLFSDRSFPAQFSQIEG-----LA 309

QY 234 ALPSKQRIQIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLSSPQELGHAFIDNV 293
 DB 310 SLGAPDEVIEKLA-----IYWFTEVFLCKEGDSIKAYGAGLLSSFGELQYCLSDRP 362

QY 294 RVLPLELDIILRPENTSTPQETLFSIRHFDLVE 328
 DB 363 KLLPLELEKTAQCEYSVTETQPLYYVAESFSDAKE 397

RESULT 18
 WHUF
 phenylalanine 4-monooxygenase (EC 1.14.16.1) - human
 N:Alternate names: phenylalanine 4-hydroxylase
 C:Species: Homo sapiens (man)
 C>Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 21-Jul-2000
 C:Accession: A00508; S02687; I52416; I54346; S74142; I54257
 R:Kwok, S.C.M.; Ledley, F.D.; Dilella, A.G.; Robson, K.J.H.; Woo, S.L.C.
 Biochemistry 24, 556-561, 1985
 A:Title: Nucleotide sequence of a full-length complementary DNA clone and amino acid seq
 A:Reference number: A00508; MUID:85199778; PMID:2986678
 A:Accession: A00508
 A:Molecule type: mRNA
 A:Residues: 1-452 <KWO>
 A:Cross-references: GB:K03020; NID:G189936; PIDN:AAA60082.1; PID:G189937
 R:Cotton, R.G.H.; McAdam, W.; Jennings, I.; Morgan, F.J.
 Biochem. J. 255, 193-196, 1988
 A:Title: A monoclonal antibody to aromatic amino acid hydroxylases. Identification of th
 A:Reference number: S02687; MUID:89061656; PMID:2481704
 A:Accession: S02687
 A:Molecule type: protein
 A:Residues: 131-144 <COF>
 R:Konecki, D.S.; Wang, Y.; Trefz, F.K.; Lichter-Konecki, U.; Woo, S.L.
 Biochemistry 31, 8363-8368, 1992
 A:Title: Structural characterization of the 5' regions of the human phenylalanine hydrox
 A:Reference number: I52416; MUID:922999453; PMID:1326329
 A:Accession: I52416
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-20 <KON>
 A:Cross-references: GB:S44225; NID:G255493
 R:Abadie, V.; Jaruzelska, J.; Lyonnet, S.; Millaesau, P.; Berthelon, M.; Rey, F.; Munni
 Hum. Mol. Genet. 2, 31-34, 1993
 A:Title: Illegitimate transcription of the phenylalanine hydroxylase gene in lymphocytes
 A:Reference number: I54346; MUID:93258345; PMID:8098245
 A:Accession: I54346
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 381-389, C' 391-405 <ABR>
 A:Cross-references: GB:S61296; NID:G300410; PIDN:AAD13926.1; PID:94261626
 R:Kowlessar, D.; Citron, B.A.; Kaufman, S.
 Arch. Biochem. Biophys. 333, 85-95, 1996

A:Title: Recombinant human phenylalanine hydroxylase: novel regulatory and structural pr
 A:Reference number: S74142; MUID:96400381; PMID:8806757
 A:Accession: S74142
 A:Molecule type: protein
 A:Residues: 2-21 <KOW>
 R:Eigel, A.; Dworniczak, B.; Kalaydjieva, L.; Horst, J.
 Hum. Genet. 87, 739-743, 1991
 A:Title: A frameshift mutation in exon 2 of the phenylalanine hydroxylase gene linked to
 A:Reference number: I54257; MUID:92039642; PMID:1682235
 A:Accession: I54257
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 53-54, 'LRRMM' <ETG>
 A:Cross-references: GB:S62592; NID:G238240; PIDN:AAB20205.1; PID:G238241
 A>Note: a mutant sequence found in patients with phenylketonuria
 C:Genetics:
 A:Gene: GDB:PAH
 A:Cross-references: GDB:119470; OMIM:261600
 A:Map position: 12q24.1-12q24.1
 A>Note: a defect in this gene can cause phenylketonuria
 C:Complex: homodimer
 C:Function:
 A:Description: catalyzes the 4'-hydroxylation of phenylalanine to tyrosine by tetrahydrot
 A:Pathway: tyrosine biosynthesis; phenylalanine catabolism
 C:Superfamily: phenylalanine 4-monooxygenase
 C:Keywords: bioprotein; homodimer; iron; metalloprotein; monooxygenase; oxidoreductase; p
 F:16/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status p
 F:23/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict
 F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.3%; Score 213.5; DB 1; Length 452;
 Best Local Similarity 26.6%; Pred. No. 1.7e-09;
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWYRLSSGRFSLWKSYPFRFLDYLEAFGLSDFLDH-----QAVIKFELETHP 155
 DB 184 KKTGTGTVFKLKSLYTHA---CYENHIFPLEKYCGFHEDNPQLQEDVVSQFLQCTGF 240

QY 156 SYYPVSGFVAPHQYLSLDQRYPTASVMRTLDKNFSLDPLDHLHGVPLWLLHPSFS 215
 DB 241 RLRPVAGLLSSRDLGLGAPRVPHCTQYIRHSGKPMYTPEDPICHLLGHVPLFSDRSEA 300

QY 216 EPPINMGRLETKVIEKQVLPKQRIQIOTLQSNLIAIVRCFWFTVESGLIENHGRKAYG 275
 DB 301 QFSQIEG-----LASLGAPDEVIEKLA-----IYWFTEVFLCKEGDSIKAYG 344

QY 276 NVLSSPQELGHAFIDNVRLPDLDDIILRPENTSTPQETLFSIRHFDLVE 328
 DB 345 AGLSSFGELQYCLSEKPLPLELEKTAQYVTETQPLYYVAESFSDAKE 397

RESULT 19
 S15758
 phenylalanine 4-monooxygenase (EC 1.14.16.1) - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 31-Mar-2000
 C:Accession: S15758; S08678
 R:Ledley, F.D.; Grenett, H.E.; Dunbar, B.S.; Woo, S.L.C.
 Biochem. J. 267, 399-406, 1990
 A:Title: Mouse phenylalanine hydroxylase. Homology and divergence from human phenylalanin
 A:Reference number: S15758; MUID:90241147; PMID:2334400
 A:Accession: S15758
 A:Molecule type: mRNA
 A:Residues: 1-453 <LED>
 A:Cross-references: EMBL:X51942; NID:G53683; PIDN:CAA36205.1; PID:G53684
 C:Superfamily: phenylalanine 4-monooxygenase
 C:Keywords: bioprotein; iron; metalloprotein; oxidoreductase; phosphoprotein
 F:285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.2%; Score 212.5; DB 2; Length 453;
 Best Local Similarity 29.9%; Pred. No. 2e-09;
 Matches 61; Conservative 35; Mismatches 87; Indels 21; Gaps 5;

QY 104 RNLM---YRLSSRFSLMKSCYCRFFLDYLEAFGLSDPLDH---QAVIKFELETHFSY 158
 DB 184 RKTWGTFRRLTKALYKTHACYEHHNIFPLLEKICGFRDNIPQLEDSVSPFQDTCTGRFLR 243
 QY 159 PVSGFVAHQYLSLQDRYFPPIASVMTLTKDNFSLTPDLIDLLGHWPLMLHPFSSEFF 218
 DB 244 PVAGLLSRDPLGLAFRRVFCQYIRHSGKPMWTEPDICHELLGHVPLPSDRSFAQFS 303
 QY 219 INMGRLFTKVIKQVALPSKKQRIQTLQSNLAIYRCFWFTVESGLIENHGRKAYGAVL 278
 DB 304 QEIG-----LASLGAPDEYIEKLTAT-----IYFTVEFGICKRGDSIYAYGAGL 347
 QY 279 ISSPOLGHAFLDNVRLVPLELDQ 302
 DB 348 ISSFGLQYCLSDKPKLLPLELEK 371

RESULT 20
 tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine

N/Alternate names: tyrosine 3-hydroxylase
 C/Species: Bos primigenius laurus (cattle)
 C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 31-Mar-2000
 C/Accession: J00039; A27687
 R/Notes: S.; Stehle, A.D.; Lamouroux, A.; Mallet, J.; Thoenen, H.
 J. Neurochem. 51, 572-578, 1988
 A/Title: Predicted amino acid sequence of bovine tyrosine hydroxylase and its similarity
 A/Reference number: J00039; MUID:88274405; PMID:2899135
 A/Accession: J00039
 A/Molecule type: mRNA
 A/Residues: 1-491 <SAA>
 A/Cross-references: GB:M36705; NID:G153796; PIDN:AAA0798.1; PID:G163797
 A/Experimental source: adrenal medulla
 A/Biochem. Biophys. Res. Commun. 151, 1446-1453, 1988
 A/Title: Characterization of the catalytic domain of bovine adrenal tyrosine hydroxylase
 A/Reference number: A27687; MUID:88183482; PMID:2895648
 A/Accession: A27687
 A/Molecule type: protein
 A/Residues: 153-158,160,162-169 <ABA>
 C/Comment: This protein is the rate-limiting enzyme in the biosynthesis of catecholamine
 C/Superfamily: phenylalanine 4-monooxygenase
 C/KeyWords: bioprotein; iron; metalloprotein; oxidoreductase; phosphoprotein
 F/2-491/Product: tyrosine 3-monooxygenase #status predicted <Mx>
 F/8/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status predicted
 F/19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status predicted
 F/40/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predicted
 F/324,329,369/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.2%; Score 211.5; DB 2; Length 491;
 Best Local Similarity 26.7%; Pred. No. 2.7e-09;
 Matches 64; Conservative 4; Mismatches 104; Indels 31; Gaps 5;

QY 107 WYRLSSRFSLMKSCYCRFFLDYLEAFGLSDPLDH-----QAVIKFELETHFSY 158
 DB 226 WKEVSTRLRGLYPTHAC---EHLEAFELLERFCGYREDRIPOLEDSVSPFLKERGFOLR 282
 QY 159 PVSGFVAHQYLSLQDRYFPPIASVMTLTKDNFSLTPDLIDLLGHWPLMLHPFSSEFF 218
 DB 283 PVAGLLSRDPLGLAFRRVFCQYIRHSGKPMWTEPDICHELLGHVPLPSDRSFAQFS 342
 QY 219 INMGRLFTKVIKQVALPSKKQRIQTLQSNLAIYRCFWFTVESGLIENHGRKAYGAVL 278
 DB 343 QDIG-----LASLGAPDEYIEKLTAT-----IYFTVEFGICKRGDSIYAYGAGL 347
 QY 279 ISSPOLGHAFLDNVRLVPLELDQIRLPNTSTPOETLPSIRNF---DELVELTSKLE 334
 DB 387 ISSYGELEHLSSEBEIRAFDPDAAVQPDQYQYVVSSEPSDAXDKLRYSASKIQ 446

RESULT 21
 J00766
 phenylalanine 4-monooxygenase (EC 1.14.16.1) - fruit fly (Drosophila melanogaster)

N/Alternate names: phenylalanine 4-hydroxylase; tryptophan hydroxylase
 C/Species: Drosophila melanogaster
 C/Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 31-Mar-2000
 C/Accession: J00766
 R/Notes: G.; Requena, J.M.; Jimenez-Ruiz, A.; Lopez, M.C.; Ugarte, M.; Alonso, C.
 gene 93, 213-219, 1990
 A/Title: Sequence and expression of the Drosophila phenylalanine hydroxylase mRNA.
 A/Reference number: J00766; MUID:9103030; PMID:2121612
 A/Accession: J00766
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-453 <MOR>
 A/Cross-references: GB:M32802; NID:G158035; PIDN:AAA69513.1; PID:G158036
 C/Genetics:
 A/Gene: FlyBase:TpH
 A/Cross-references: FlyBase:FBgn0005770
 C/Superfamily: phenylalanine 4-monooxygenase
 C/KeyWords: bioprotein; iron; metalloprotein; monooxygenase; oxidoreductase; phenylke-
 F/285,290,330/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.0%; Score 207; DB 2; Length 453;
 Best Local Similarity 27.9%; Pred. No. 5.7e-09;
 Matches 75; Conservative 39; Mismatches 109; Indels 46; Gaps 10;

QY 107 WYRLSSRFSLMKSCYCRFFLDYLEAFGLSD---PLDH-----QAVIKFELETHFSY 158
 DB 187 WGIIFRRLTKLYTHACR---RYNHVPLVDNCGFRDNIPQLEDSVSPFLRDCGTFLR 243
 QY 159 PVSGFVAHQYLSLQDRYFPPIASVMTLTKDNFSLTPDLIDLLGHWPLMLHPFSSEFF 218
 DB 244 PVAGLLSRDPLGLAFRRVFCQYIRHSGKPMWTEPDICHELLGHVPLPSDRSFAQFS 303
 QY 219 INMGRLFTKVIKQVALPSKKQRIQTLQSNLAIYRCFWFTVESGLIENHGRKAYGAVL 278
 DB 304 QEIG-----LASLGAPDEYIEKLTAT-----IYFTVEFGICKRGDSIYAYGAGL 347
 QY 279 ISSPOLGHAFLDNVRLVPLELDQIRLPNTSTPOETLPSIRNFVELTSKLEMD 338
 DB 348 ISSYGELEHLSSEBEIRAFDPDAAVQPDQYQYVVSSEPSDAXDKLRYSASKIQ 446
 QY 339 QGV---LESTP-----LYNDEKLSGFEVL 360
 DB 398 KTIKFNISIRPGVKN---ATQSVLEV 424

RESULT 22

hypothetical protein ZK1290.2 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T34509
 R/Title: A.
 submitted to the EMBL Data Library, July 1995
 A/Description: The sequence of C. elegans cosmid ZK1290.
 A/Reference number: Z21535
 A/Accession: T34509
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-575 <TAI>
 A/Cross-references: EMBL:U21308; PIDN:AA93319.1; GSPDB:GN00020; CESP:ZK1290.2
 A/Experimental source: strain Bristol N2; clone ZK1290
 C/Genetics:
 A/Gene: CESP:ZK1290.2
 A/Map position: 2
 A/Insertion: 49/3; 72/2; 105/3; 223/1; 257/3; 280/2; 310/2; 346/1; 391/2; 433/3; 523/2

Query Match 10.9%; Score 206.5; DB 2; Length 575;
 Best Local Similarity 25.2%; Pred. No. 8.5e-09;
 Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNLM---YRLSSRFSLMKSCYCRFFLDYLEAFGLSDPLDH-----QAVIKFELETHFSY 155
 DB 294 RKTWGTFRRLTKALYKTHACYEHHNIFPLLEKICGFRDNIPQLEDSVSPFLKERGFOLR 243

QY 156 SYYPVSGFVAPHOYLSLLODRYFFIASVMRTLDKONFSLTDPDLHLLGHVPWLLHPSFS 215
DB 351 RVRPVAGYLSARDFLAGLAVRYFFCTQYVRHADPFYTPEDTVHLMGCHMALFADPDFA 410
QY 216 EPIFMNGRLFTKVKVQALPSKQRIQTLQSNLAIIVRCFWFTVBSGLI----- 265
DB 411 QFSQBIG-----LASLGASEEDLKKLATL-----YFISIERGLSSDDAADSVPK 454
QY 266 ---ENHEGRKAYGAVLSSQELGHAFIDNVRLPLELDQIIRLPNTSTQETLSIRH 322
DB 455 ENGSNHERFKYAGLSSAGELQHAVESATIRFDRVVEQECLITTFQAYFYTN 514
QY 323 PDELVELTSKLEWMLDOGLLESIPLYNQEKYLSGFEVL 360
DB 515 FEEAQ-----KLMTNNKRPFFVRIN--PYTESVEVL 547

RESULT 23

A42271

tryptophan 5-monooxygenase (EC 1.14.16.4) - fruit fly (*Drosophila* sp.)

N/Alternate names: tryptophan hydroxylase

C/Species: *Drosophila* sp.

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000

C/Accession: A42271

R/Neckameyer, W.S.; White, K.

J. Biol. Chem. 267, 4199-4206, 1992

A/Title: A single locus encodes both phenylalanine hydroxylase and tryptophan hydroxylase

A/Reference number: A42271; MUID: 92156168; PMID: 1371286

A/Accession: A42271

A/Molecule type: mRNA

A/Residues: 1-453 <NC>

A/Note: sequence extracted from NCBI backbone (NCBIN:82902, NCBI:82903)

C/Genetics:

A/Gene: FlyBase:Tph

A/Cross-references: FlyBase:FBgn0005770

C/Superfamily: phenylalanine 4-monooxygenase

C/Keywords: biotin; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein

F:284,289,329/Binding site: iron (His, Glu) #status predicted

Query Match 10.4%; Score 197; DB 1; Length 453;

Best Local Similarity 27.5%; Pred. No. 3.6e-08;

Matches 74; Conservative 38; Mismatches 111; Indels 46; Gaps 10;

QY 107 WYRLSSRFSLWKSYPFRFLDYLEAFGLSD---FLDH---QAVIKFLETHFSY 158
DB 186 WGIIFRNLTKLYKTHACR---EYNHVPLLVNDCGFREDNIPQLEDVSNFLRDTGFTLR 242
QY 159 PVSGFVAPHOYLSLLODRYFFIASVMRTLDKONFSLTDPDLHLLGHVPWLLHPSFS 218
DB 243 PVAGLLSSRDPLAGLAFRVFHSQYIRHPSKPMYTPEDVCHLMGHVPLFADPAFAQS 302
QY 219 INMGRLFTKVKVQALPSKQRIQTLQSNLAIIVRCFWFTVBSGLIENHEGRKAYGAVL 278
DB 303 QEIG-----LASLGAPDDYIEKLSL-----IFWTFVEYGLCRQSGELKAYGAGL 346
QY 279 ISSQELGHAFIDNVRLPLELDQIIRLPNTSTQETLSIRHFDDELVELTSKLEWMLD 338
DB 347 LSSYGELEYCLTKPKQDFE-----PEVTGV---TKYPIQFQPLYVADSFFETAKE 396

QY 339 QGL--LESIP-----LYNQEKYLSGFEVL 360

DB 397 KTIKFANSIPRPFGRVYN--AYTQSVLEVL 423

RESULT 24

JC4888

phenylalanine 4-monooxygenase (EC 1.14.16.1) - fruit fly (*Drosophila melanogaster*)

N/Alternate names: phenylalaninase; phenylalanine 4-hydroxylase

C/Species: *Drosophila melanogaster*

C/Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 20-Jun-2000

C/Accession: JC4888

R/Ruiz-Vazquez, P.; Moulard, M.; Silva, F.J.

Biochem. Biophys. Res. Commun. 225, 238-242, 1996
A/Title: Structure of the phenylalanine hydroxylase gene in *Drosophila melanogaster* and
A/Reference number: JC4888; MUID: 96332435; PMID: 8769124
A/Accession: JC4888

A/Molecule type: mRNA

A/Residues: 1-452 <RUI>

A/Cross-references: EMBL:X98116; NID: g1483312; PIDN: CAA66797.1; PID: g1483313

A/Note: the translation of the nucleotide sequence is not complete in this paper

C/Comment: This enzyme is a member of the pterin-dependent aromatic amino acid hydroxylase

C/Genetics:

A/Gene: Pah

A/Introns: 10/3; 235/1; 322/3; 399/2

C/Superfamily: phenylalanine 4-monooxygenase

C/Keywords: biotin; iron; metalloprotein; oxidoreductase

F:284,289,329/Binding site: iron (His, Glu) #status predicted

Query Match 10.3%; Score 194; DB 2; Length 452;

Best Local Similarity 27.1%; Pred. No. 6.3e-08;

Matches 73; Conservative 39; Mismatches 111; Indels 46; Gaps 10;

QY 107 WYRLSSRFSLWKSYPFRFLDYLEAFGLSD---FLDH---QAVIKFLETHFSY 158
DB 186 WGIIFRNLTKLYKTHACR---EYNHVPLLVNDCGFREDNIPQLEDVSNFLRDTGFTLR 242
QY 159 PVSGFVAPHOYLSLLODRYFFIASVMRTLDKONFSLTDPDLHLLGHVPWLLHPSFS 218
DB 243 PVAGLLSSRDPLAGLAFRVFHSQYIRHPSKPMYTPEDVCHLMGHVPLFADPAFAQS 302
QY 219 INMGRLFTKVKVQALPSKQRIQTLQSNLAIIVRCFWFTVBSGLIENHEGRKAYGAVL 278
DB 303 QEIG-----LASLGAPDDYIEKLSL-----IFWTFVEYGLCRQSGELKAYGAGL 346
QY 279 ISSQELGHAFIDNVRLPLELDQIIRLPNTSTQETLSIRHFDDELVELTSKLEWMLD 338
DB 347 LSSYGELEYCLTKPKQDFE-----PEVTGV---TKYPIQFQPLYVADSFFETAKE 396

QY 339 QGL--LESIP-----LYNQEKYLSGFEVL 360

DB 397 KTIKFANSIPRPFGRVYN--AYTQSVLEVL 423

RESULT 25

CB7449

phenylalanine-4-hydroxylase [imported] - *Caulobacter crescentus*C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: CB7449

R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

n.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A87249; MUID: 21173698; PMID: 11259647

A/Accession: CB7449

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-294 <STO>

A/Cross-references: GB:AE005673; NID: g13423011; PIDN: AAK23591.1; GSPDB: GN00148

C/Genetics:

A/Gene: CCI612

Query Match 10.0%; Score 189.5; DB 2; Length 294;

Best Local Similarity 24.1%; Pred. No. 8.2e-08;

Matches 63; Conservative 43; Mismatches 108; Indels 47; Gaps 6;

QY 118 WKSYPFRFLDYLEAFGLSDFLDHOAVIKFPE---LETHFSYYP----- 159
DB 24 WETTYQASHDWITLYERQDTMLHGRACDEFMRGLDALDLHRSIGDPFARINEELKRLTG 83
QY 160 -----VSGFVAPHOYLSLLODRYFFIASVMRTLDKONFSLTDPDLHLLGHVPWLLHPSF 214
DB 84 WTVVAVPGLVDDVFDFDLNRRFPAGQIRKPHLDYLOEPDIFHDVFGHVPMLTDPVF 143

QY 215 SEFFINMRLEFTKVIKVKQALPSKKORIQTLOSNIILAVRCFWTVESGLIENHGRKAY 274
Db 144 ADY-----MAYGEGRRALGL-GRLLNLRKLYWTVFGLNMTFAGLRIT 188
QY 275 GAVLISSPOELGHAFFID-NKRVLPLELDQIRLPFTSTPQSTLSIHAFELVELT--- 330
Db 189 GAGIVSSRTSISIFALDDPSIRIGFDLERVMTLYRIDDFQOYFVIDSIQTLEVTLRD 248
QY 331 -----SKLEWMLDOGLLESIP 346
Db 249 FGAIVYERLASVSDIGVAEIVP 269

RESULT 26

T25453
tyrosine 3-monooxygenase (EC 1.14.16.2) B0432.5 [similarity] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_rev15ion 15-Oct-1999 #text_change 19-May-2000
C:Accession: T25453
R:Henkhaus, J.; Wohldmann, P.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid B0432.
A:Reference number: Z20038
A:Accession: T25453
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <HFN>
A:Cross-references: EMBL:U80836; PIDN:AA837888.1; GSPDB:GN00020; CESP:B0432.5
A:Experimental source: strain Bristol N2; clone B0432
C:Genetics:
A:Gene: CESP:B0432.5
A:Map position: 2
A:Introns: 117/2; 183/1; 273/2; 302/3; 369/3
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioprotein; oxidoreductase

Query Match 9.4%; Score 178.5; DB 2; Length 404;
Best Local Similarity 24.3%; Pred. No. 9.6e-07;
Matches 77; Conservative 41; Mismatches 114; Indels 85; Gaps 11;

QY 19 KLRQSLSPFQNSQSLORAVSTPVSRYRIILCKENKQA-LARH-----KCTISILE 69
Db 32 QLIHSAELTLQNHVAL-----TKFSIFAKKLSDEKQSGIWPFRHISLQCSKCTIKYE 86
QY 70 FPKMLLPVHLISLKNQREGSTDMVAVSTPPFNRLMWRLSSRFSLSMKSCYCPRPFLDY 129
Db 87 -----PTTPRHGHGHDVAVIARKTLD--QALEFRFGDEIGY-----VDY 126
QY 130 LE-----AFGLSDP-LDHQAVI-----KFELETH 154
Db 127 TEERHATWKAIVTEKLDGLHSHTCAYRQNLKLDSEKVLTDRIIPQIRDVNKKFLQCKTG 186
QY 155 FSYVPVSGFVAPHOYLSLQDRYPPIASVWRTLDKNFSLPTDLIHLLGHVPWMLHPSP 214
Db 187 FELRCGSLLSARDPLASLAFRVFQTTTYLRHKKSPHNSPEPDLLHLLGHVPWMSDFLL 246
QY 215 SEFFINMRLEFTKVIKVKQALPSKKORIQTLOSNIILAVRCFWTVESGLIENHGRKAY 274
Db 247 AQMSQDIG-----LMSLIGASDEHIKLSF-----VVMFVFERGLCKEDGKKAII 290
QY 275 GAVLISSPOELGHAFFID 291
Db 291 GAGIISAVGEIMHACSD 307

RESULT 27

T23494
phenylalanine 4-monooxygenase (EC 1.14.16.1) K0898.4 [similarity] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_rev15ion 15-Oct-1999 #text_change 19-May-2000
C:Accession: T23494
R:Smey, R.
submitted to the EMBL Data Library, October 1995

A:Reference number: Z19747
A:Accession: T23494
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-457 <WLL>
A:Cross-references: EMBL:Z66497; PIDN:CAA91286.1; GSPDB:GN00020; CESP:K0898.4
A:Experimental source: clone K0898
C:Genetics:
A:Gene: CESP:K0898.4
A:Map position: 2
A:Introns: 12/3; 51/3; 116/1; 170/2; 236/1; 281/2; 323/3; 400/2
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioprotein; oxidoreductase

Query Match 9.3%; Score 175.5; DB 2; Length 457;
Best Local Similarity 25.5%; Pred. No. 2e-06;
Matches 56; Conservative 35; Mismatches 80; Indels 49; Gaps 5;

QY 143 QAVIKFELETHPSYVPGFVAPHOYLSLQDRYPPIASVWRTLDKNFSLTPDLIDH 202
Db 228 QDVSDFLKQCTGYTIRPVAAGLSSRDPLAGLAFVHFSTGYIRHSAKPYTPEPDLCHL 287
QY 203 LGHVPLMLHPSPSEFFINMRLEFTKVIKVKQALPSKKORIQTLOSNIILAVRCFW 258
Db 288 LGHVPLFADVERFQPSQETIGLASGAPDVIETKATL-----YVF 327
QY 259 TVESGLIENHGRKAYGAVLISSPOELGHAFFIDNVRVLPLE-LDQIRLPFTSTPQST 316
Db 328 TIERGICQODGEKKAAGLSSRFELOYLSLSDPEVDDPAVCCVYTPITTEYQPKYF 387
QY 317 L-----FSIRH--FDELVELTSKL 333
Db 388 LAESFASAKNKKLSMAATINRPFQIRYVAYQREVLIDKY 427

RESULT 28

T45787
disease resistance-like protein - Arabidopsis thaliana
N:Alternate names: protein F26013.200
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_rev15ion 04-Feb-2000 #text_change 04-Feb-2000
R:Delany, M.; Berger, C.; Cooke, R.; Grellet, F.; Laurie, M.; Mewes, H.W.; Lemcke, T.
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23013
A:Accession: T45787
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1253
A:Cross-references: EMBL:AL133452
A:Experimental source: cultivar Columbia; BAC clone F26013
C:Genetics:
A:Map position: 3
A:Introns: 138/2; 490/3; 589/3; 1047/3
A:Note: F26013.200

Query Match 5.8%; Score 109.5; DB 2; Length 1253;
Best Local Similarity 20.3%; Pred. No. 1.6;
Matches 81; Conservative 61; Mismatches 130; Indels 127; Gaps 18;

QY 77 VHLISLSTKNQREGCS-----TDMAVSTPFFRNLMWRLSS-----RF 115
Db 421 IHL-----EGCGFPFVEINVLVEKCLVSMN-ESRVVMHNLIGISGRKIINGKRRS 472
QY 116 SLMKSCYCPRPFLDYLEARG-----LSDFLDHQA-----KFFEL-----ET 153
Db 473 RUKPKLIRYFLEDRQVGSBDEIALFIDPEALSFDVNPVAFEMWYNLAKICSSNGN 532
QY 154 HPSYVPGFVAPHOYLSLQDRYPPIASVWRTLDKNFSL-----TPD-- 197
Db 533 HVALHPKGVKSPEBLRLHWHFPLSLDQDNTNRLVILMCKSKQLRIMBGTKEG 592
QY 198 -LTHDLLGHVPWMLHPSPSEFFINM-----GRL-----FTYVIE-----KVQ 233

Db 1143 -----QQVHYNQ 1150

RESULT 30

A40996
phenylalanine 4-monooxygenase (EC 1.14.16.1) - Chromobacterium violaceum
N/Alternate names: phenylalaninase; phenylalanine 4-hydroxylase
C/Species: Chromobacterium violaceum
C/Date: 03-Aug-1992 #sequence_revision 03-Aug-1992 #text_change 31-Mar-2000
C/Accession: A40996; B40996
R/Onishi, A.; Liotta, L.J.; Benkovic, S.J.
J. Biol. Chem. 266, 18454-18459, 1991
A/R>Title: Cloning and expression of Chromobacterium violaceum phenylalanine hydroxylase
A/Reference number: A40996; PMID:92011593; PMID:1655752
A/Accession: A40996
A/Molecule type: DNA
A/Residues: 1-296 <ONI>
A/Cross-references: GB:M55915; NID:G144481; PIDN:AAA23115.1; PID:G144482
A/Accession: B40996
A/Molecule type: protein
A/Residues: 1-20 <ONS>
A/Experimental source: strain ATCC 12540
C/Keywords: bioplerin; oxidoreductase

Query Match 5.6%; Score 106.5; DB 2; Length 296;
Rest Local Similarity 27.3%; Pred. No. 0.4;
Matches 27; Conservative 22; Mismatches 49; Indels 1; Gaps 1;

QY 126 FLDYLEAFGLSDLEL-DHQAVIKFFLETHFSYYPVSGFVAPHQVLSLLQDRVPEPIASVM 184
DB 63 FLEGLELEVADRPDENKLEKLMATCWKIVAVPGLIPDDVFFEHANREFVPTWWL 122

QY 185 RTLDKDNLSLPDLIHLLGHVFWLLHPSFSBEFFNMGR 223
DB 123 REPHQDLYQEDVDVDFHDLFGHVELLLNPFVADYLEAYGK 161

RESULT 31

S44054
genome polyprotein - Marburg virus (strain Popp)
N/Alternate names: structural protein L
N/Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C/Species: Marburg virus
A/Variety: strain Popp
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
C/Accession: S44054; S32776
R/Bukreyev, A.A.; Volchkov, V.E.; Blinov, V.M.; Dryga, S.A.; Netesov, S.V.
submitted to the EMBL Data Library, January 1994
A/Description: Full-length nucleotide sequence of Marburg virus Popp strain: The comparison
A/Reference number: S44049
A/Accession: S44054
A/Molecule type: genomic RNA
A/Residues: 1-2331 <BUK>
A/Cross-references: EMBL:Z29337; NID:G450908; PIDN:CAA82542.1; PID:G450915
A/Experimental source: strain Popp
R/Bukreyev, A.A.; Netesov, S.V.
submitted to the EMBL Data Library, September 1992
A/Description: The partial nucleotide sequence of Marburg virus genome.
A/Reference number: S32775
A/Accession: S32776
A/Molecule type: genomic RNA
A/Residues: 1-2331 <BUW>
A/Cross-references: EMBL:X69494; NID:G296962; PIDN:CAA448508.1; PID:G296963
A/Experimental source: strain Popp
C/Genetics:
A/Genes: L
C/Superfamily: parainfluenza virus RNA-directed RNA polymerase
C/Keywords: ATP; nucleotidyltransferase

Query Match 5.5%; Score 104.5; DB 2; Length 2331;
Best Local Similarity 23.3%; Pred. No. 9.1;
Matches 59; Conservative 42; Mismatches 117; Indels 35; Gaps 7;

QY 11 KYIKALKLNOSLSLFONOSLOPASTVSYRRILOKENEKQALAHKICISILF 70
 DB 356 KTIIOGKLTPOQYCELF-----SLQKMGHPVLIVDALDKYKAQASVKILKPKVMEET 410
 QY 71 KQULFEVHLISLKNOREGCTDMAVAVSTPFENRNLWYLLSSRFSLMKSCPRFLDYL 130
 DB 411 FGVKFIYAKOHVHSSQSWYKTMTHLTPYLRQHIIVSNSPFSQAEIYQOHLMEWYFVZHE 470
 QY 131 EAFGL-----LSDFLDHQAIVKPELETHFSYRPGVAPRHOVYSLQDRYFPLASVNR 185
 DB 471 PLFSTKIISLDSIFIKORATVNOECWDSVFDSDVAGVNPVRFGSKVPEPF----- 523
 QY 186 TIDKDNFSITPDILHDLGHVPMILHSPSEPF-----INMGELFKVIEKQALPSK 238
 DB 524 -LGQADFSL--NQILDFEKLEYLA-PSYRNFSPSLKEKELNIGRTGK-----LPIR 572
 QY 239 KQRIOTLOSNIUA 251
 DB 573 VRRVQTLAALIA 585

RESULT 32

T38766
 probable transcription regulatory protein - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Aug-2002
 C/Accession: T38766
 R/Gentles, S.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, January 1996
 A/Reference number: Z21810
 A/Accession: T38766
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-563 <GEN>
 A/Cross-references: EMBL:Z69086; PIDN:CAA93165.1; GSPDB:GN00066; SPDB:SPAC3H8.08C
 A/Experimental source: strain 972h-, cosmid c3H8
 C/Genetics:
 A/Gene: SPDB:SPAC3H8.08C
 A/Map position: 1
 C/Supplementary: transcription regulator SPEC16G5.17, GAL4 zinc binuclear cluster homology F.14-50/Domain: GAL4 zinc binuclear cluster homology <G4>

Query Match 5.4%; Score 102; DB 2; Length 563;

Best Local Similarity 21.1%; Pred. No. 2.2; Mismatches 142; Indels 102; Gaps 16;

Matches 80; Conservative 56; Mismatches 142; Indels 102; Gaps 16;

QY 7 TIDPKYIKALKLNOSLSLFONOSLOPASTVSYRRILOKENEKQALAHKICIS 66
 DB 185 TLPIPIAATIQSLDPVILNFMVNSAGI-----TPLESSRLNLKLNIEISDEQYGHCLP 239
 QY 67 ILEFFKNLLFVHLISLKNOREGCTDM-----AVSTPFENRNLWYLLSSRFSLMKS 120
 DB 240 DKEIIOMLLBRAYATKFRIRGVNTDLCSIHVSTLVPLF--GVTEKIGKNTSDLMFA 297
 QY 121 YCPREFLDVLEAFGLSDFLDHQAIVKPELETHFSYRPGVAPRHOVYSLQDRYFPL 180
 DB 298 LC-----EIDGL-----ECVLKY--RPFPIQHPYGLKRP----- 325
 QY 181 ASVMRTLDKDNFSITPDILHDLGHVPMILHSPSEPFINMGRLPFKVIKQALPSKKQ 240
 DB 326 ---LACFPNDLSYN--FHLLGR---LDDCGVSYKSHSLTYSKFDKLESYESGLS 376
 QY 241 RI-----QTQSMLAIYRCFW-----FVYESGLINHEGKRAYAV 277
 DB 377 LILVLDIAKFYDPSNEDIQFYITLKNVFWTARVNLKOCFTILDSILDEDE-----T 429
 QY 278 LISSPQELGHAFFIDNVRVLPLELDQIIRLPNTSTPOETLFSIRHFDLVELTSLKLEML 337
 DB 430 IIGN--LGESCIQCVRLL--ISQITIL-----EKRGMVLVALLETIHIALMLAACFR 476
 QY 338 DQGL-----LESIPLYNOEK 352

DB 477 DKGEVPSDLGDTITLYVOER 496

RESULT 33

C81624
 exodeoxyribonuclease V, beta chain, probable CP0007 (imported) - Chlamydomonas pneum.
 C/Species: Chlamydomonas pneumoniae
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C/Accession: C81624
 R/Reed, T.D.; Brubham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hick
 C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Koloney, J.; McClarty, G.; Salz
 Nucleic Acids Res. 28, 1397-1406, 2000
 A/Title: Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR3
 A/Reference number: AB1500; MUID:20150255; PMID:10684935
 A/Accession: C81624
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1050 <REA>
 A/Cross-references: GB:AE002164; GB:AE002161; NID:G7188939; PIDN:AA937903.1; PID:G71
 A/Experimental source: strain AR39, HL cells
 C/Genetics:
 A/Gene: CP0007

Query Match 5.4%; Score 102; DB 2; Length 1050;

Best Local Similarity 22.3%; Pred. No. 5; Mismatches 78; Conservative 51; Mismatches 116; Indels 104; Gaps 19;

QY 46 RILOKENEKQALAHKICISIL--EPPKILF--VHLISLKNOREGCTDMA----- 95
 DB 137 RLHKKKALHNSQLVHHITNYLKODLWKVLFQEOFHLLAVRYNTSKHTSLVDKLLA 196
 QY 96 ----VSTPFENRNLWY--RLSSRFSLMKSCYCPFLDYLEA-----FG 134
 DB 197 SYTPISSYSSRVERLEQISLMHQIYNSLEIRK---QVFLDQTLAHISGKKQPS 252
 QY 135 LLSDFLDHQAIVKPELETH--FSYRPGV--GFAVAPRHOVYSLQDRYFPLASVNR----- 182
 DB 253 ILDDL--HNPVDLYNSEHSLSPFKIATFPKRLA-----RYKCAAFVLENNIS 305
 QY 183 -WRTLDKDNF-----SLTPDILHDL--LGHVPMILHSPSEPFINMGRLFKVIKQALPS 235
 DB 306 WKEHLEFCHLDIFMTILVLDQELKQNYTF--LSDESVF-----ALEKLLSS 354
 QY 236 PSKKQRIOTLOSNIUAIVRCFWFVYESGLINHEGKRAYAVLISPOELG--AFIDN 292
 DB 355 SEAPVVOAL-----RQYQVLVIDEPDIDKQWISFSN 389
 QY 293 VRVPLELDQIIRLPNTSTPOETLFSIRHFDLVELTSLKLEMLDQGL 341
 DB 390 LFISPKFTGSL-----FLIDGPKOSIYEWRSADLPYTLTAKSSPSEDKOL 434

RESULT 34

S38160
 NUP133 protein - yeast (Saccharomyces cerevisiae)
 A/Alternate names: protein YKR082w; protein YKR402
 C/Species: Saccharomyces cerevisiae
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
 C/Accession: S38160; S42011; S51915; S39123
 R/Baladron, V.; Ballester, J.P.G.; Bou, G.; del Rey, F.; Estebar, P.F.; Garcia-Cantale
 submitted to the Protein Sequence Database, March 1994
 A/Reference number: S38158
 A/Accession: S38160
 A/Molecule type: DNA
 A/Residues: 1-1157 <BAL>
 A/Cross-references: EMBL:Z28107; NID:G486562; PID:G486563; MIPS:YKR082w
 A/Experimental source: strain S288C
 R/Garcia-Cantalejo, J.; Baladron, V.; Estebar, P.F.; Santor, M.A.; Bou, G.; Remacha,
 Yeast 10, 231-245, 1994
 A/Title: The complete sequence of an 18,002 bp segment of Saccharomyces cerevisiae ci
 A/Reference number: S42009; MUID:94262327; PMID:8203164
 A/Accession: S42011
 A/Molecule type: DNA

A:Residues: 1-1157 <GAR>
A:Cross-references: EMBL:Z27116; NID:g415902; PIDN:CAA81633.1; PID:g415902
A:Experimental source: strain S288C
R:Doye, V.; Wepf, R.; Hurt, E.C.
EMBO J. 13, 6062-6075, 1994
A:Title: A novel nuclear pore protein Nup133p with distinct roles in poly(A) (+) RNA tran
A:Reference number: S51915; MUID:95112817; PMID:7813444
A:Accession: S51915
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1157 <DOY>
A:Cross-references: EMBL:X80066; NID:g624033; PIDN:CAA56372.1; PID:g624034
C:Genetics:
A:Gene: SGD:NUP133
A:Cross-references: SGD:S0001790; MIPS:YKR082w
A:Map position: 11R
C:Keywords: nucleus; transmembrane protein
F:413-423/Domain: transmembrane #status predicted <TMI>
F:641-657/Domain: transmembrane #status predicted <TM2>

Query Match 5.3%; Score 100.5; DB 2; Length 1157;
Best Local Similarity 21.4%; Pred. No. 7.5;
Matches 72; Conservative 52; Mismatches 130; Indels 83; Gaps 14;

QY 30 NSQSLQRAYSTPYSYRIILQENKEQALARKHCISILEFFKNL--LFVHLLSLSKNQR 87
DB 773 NFINNNLYDDNHLDMNHVLCCKVNKEQ-----CIQAEFYKDLGLVQTLQTLQDND- 825
QY 88 EGCSTDMAVSTPFPFNRLMYRLSSRFLSKWSCYCPFFLDYLEAFGLLSDFL-----DH 142
DB 826 ---STVSLEYET-FFNE-----FPKEFSF-----TLFEYLKHKKLANDLIFRPPQOH 868
QY 143 QAVIKFTELETHESYYPVSGFVAPHQYLSLLODRVFPIASVMRTLDKDNFSLTPDLIHL 202
DB 869 DVLQFOESA-----PKYGHVAVIQ--QILDSVADAMNTKNIITVDDSKGESI----- 917
QY 203 LQHPVLLHLSFSEFFINMGRLETKVIEKQALPSKKORIQTLQSNLIAIVRCFWFTVES 262
DB 918 -----SECELHNVAKGSLSLVEK-----DNLQINTLRKIQYNLQTLIDA 956
QY 263 GLIENHEGRKAYGAVLISSPQELGHAFIDNVRVPLELDQIIRLPFNTPQETLSIRH 322
DB 957 E--KNISNKLKGGVQICKRFNGS-----IREVFNILVEELKSTTVN 998
QY 323 FDELVELTSKLEWMLDOGLLESIPLYNOEKYLSGFEV 359
DB 999 LSDLVELYSLMD--DEESLFIPLRLSLSDGNLLNPEV 1033

RESULT 35
T09220
exocyst complex protein sec5 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09220
R:Kee, Y.; Yoo, J.S.; Hazuka, C.D.; Peterson, K.E.; Hsu, S.C.; Scheller, R.H.
Proc. Natl. Acad. Sci. U.S.A. 94, 14438-14443, 1997
A:Title: Submit structure of the mammalian exocyst complex.
A:Reference number: Z16617; MUID:98070770; PMID:9405631
A:Accession: T09220
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-924 <KEE>
A:Cross-references: EMBL:AF032666; NID:g2827157; PIDN:AAC01578.1; PID:g2827158
A:Experimental source: tissue type brain
C:Genetics:
A:Gene: sec5

Query Match 5.2%; Score 99; DB 2; Length 924;
Best Local Similarity 19.9%; Pred. No. 7.4;
Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;

QY 19 KLRQSLSLFFQNSQLQRAYSTPYSYRIILQENKEQALARKHCISILEFFKNLFFVH 78

DB 233 KVEGSMQTKLEN--VLNRASNTADTLFOEVLGRKDK--ADSTRNALNVLOREFK-FLFNL 286
QY 79 LLSLSKNQREGCSTDMAVVSTPFPFNRLMYRLSSRF-----SLWKSYPFRFFLDYLEAF 133
DB 287 PLNTKRNIQKQ--DYDVVIND-----YEKAKSLFGKTEVQVFKY-----YAEVE 329
QY 134 GLSDFLDHOAVIKFPELETHESYYPVSGFV-----APHOY-LSLLQ 174
DB 330 AGIEDL--RELLKKU-LETPSTLHDQKRYIRYLSLDLHAPGCPAWQCIGAHKWTLLKMQ 386
QY 175 D-----RYFIASVMRTLDKDNFSLTPDLIHLGLHV----- 206
DB 387 DCKEGHMKSLKGNPGPHSPM-----LGLDN-----DARPSVLGHLQSOTASLKRGSFQS 435
QY 207 -----PWLKH-----PSSEFFINM--GRLETKVIEK--VQALPSK 238
DB 436 GRDQTRWYKTPHRVAFVEKLTKLVLSQLPFWKLMISYVNGSLFSETAESKQIERSKNV 495
QY 239 KOR-----IOTLQSNLIAIVR--CFWFTVESGLIENHEGRKAYGAVLIS--SPQELG 286
DB 496 RQRQNDFKMIQOEVMHSLVKLIRGALLPFSREG-----DGRQ-YGGEVQVQAEISGQWLA 549
QY 287 HAFIDNVRV-----LPLELDQIIRLPFNTPQETLSIR-----HFDLV 327
DB 550 HV-IQTIRLTVESLTALTEIPNDMLQII-----QDLILDRLIHCIMVTLQHTAAEIK 599
QY 328 ELTSKLEWMLDOGLLESIPLYNOEKYLSGFEV 360
DB 600 RLAEKEDWIDNEGLTSLPCQFEQSIVHLSQLS 632

RESULT 36
D81265
Hypothetical protein Cj1679 [imported] - Campylobacter jejuni (strain NCTC 11168)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: D81265
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-584 <PAR>
A:Cross-references: GB:AL139079; GB:AL111168; NID:g6968971; PIDN:CAB73666.1; PID:g696909;
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1679

Query Match 5.2%; Score 98.5; DB 2; Length 584;
Best Local Similarity 20.2%; Pred. No. 4.4;
Matches 81; Conservative 67; Mismatches 154; Indels 99; Gaps 19;

QY 10 PKYILKALKRQSLSLFFQNSQLQRAYSTPYSYRIILQENKEQALARKHCISIL 68
DB 29 PRFALEFATISSVKYLSLFOKALYQAEIPLNPTSFNGLMIAKSYIEN--LRLEALNLL 86
QY 69 -----BFFKNLLFVHLLSLSKNQRGCGSTDMAVVSTPFPFNRLMYRLSSRFS 116
DB 87 QTLTRKDDLEDELKLELAFIYKLSNKEESEQIFKEL--LSKDMYNLN----- 133
QY 117 LKWSYCPFRFFL-DYLEAFG-----LLSDFLDHOAVIKFPELETHESYYPVSGFVAPHQ 168
DB 134 LKQVYAEIYFKHDTKALNAHEHLCHFMDLID-----KLQ-----KGIIAEQT 177
QY 169 YLSL--LQDRYFPIASVMRTLDKDNFSLT---PDLIHLGLHVPLLHPSF----- 214
DB 178 NLNLVKLEDR-----LHSKTKENLTISKIEDFLTHQILPQKAYLLFKLFRISSLSLELF 230
QY 215 -----SEFFINMGRLETKVIE---KVALPSKKQRIQTLQSNLIAIVRCFWFTVE 261

```

Db 231 OSLOANOHQFMON---YAKVLEFNSNYCAHYAAKCLISDSHTYQEDLAVLLMR 286
Oy 262 SGLIEN-HEGRKAYGAVLISSPOLGHAFLDNRVPLP- --LDQIR- --LPRNTSTPQ 314
Db 287 MGVDDNFEKGKKYKESRLEFAAHNETFSTYHYNESLAENKFKGVDAFKKKEVAVFCEOGG 346
Oy 315 ETLFSIRHDELVELTSKLEWMDQGLLESIPLYNOEKYLS 355
Db 347 DTIMYARCLKICKIKASVLPFAPOSAMYEMFK--NOIKFELN 385

```

RESULT 37

```

C64208
Hypothetical protein MG075 - Mycoplasma genitalium
C/Species: Mycoplasma genitalium
C/Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 07-Dec-1999
C/Accession: C64208
R/Praser: C.M. Cocayne, J.D. White, O. Adams, M.D. Clayton, R.A. Fleischmann, R.D. M. Finkham, J. Nguyen, D. Utecherback, T.R. Saudek, D.M. Phillips, C.A. Merrick, J. Science 270, 397-403, 1995
A>Title: The minimal gene complement of Mycoplasma genitalium.
A/Accession: C64208
A/Reference number: A64200; MUID:96026346; PMID:7569993
A/Molecule type: DNA
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Residues: 1-1024 <TIGR>
A/Cross-references: GB:U39687; GB:U43967; NID:G1045744; PID:G1045751; TIGR:MG075
A/Experimental source: strain G-37
A/Genetics:
A/Genetic code: SGC3

```

```

Query Match 5.2%; Score 98.5; DB 2; Length 1024;
Best Local Similarity 19.9%; Pred. No. 9.3;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

```

```

Oy 29 ONSLOKAVSPSYRYIIQKENEK---QALARKKISLIEFFKILLFVHLLSKN 85
Db 564 QOTDLSKLVFV---IGDILSETNVKIKTLHAVKNELLSLVERASTLKIKHL----- 613
Oy 86 QREGGSDMAVSTPFRRNLMVLLSSRSIMKSYCPRFPIIDYEAFGLSDFLDHQA 145
Db 614 -----NVQKVLVDKELKNSFIK-----ELNFPDITKDIPTT 647
Oy 146 IR--FPELETH-----ESYRVSGFVAPHOYLSDRFPPLASVAKTLDQNF 192
Db 648 IKKVLFESENKTKLRKKEENGEFGYMAKFIYGFPSAENTFSAI-----DKT 698
Oy 193 SLTPDLIHLGLHVPWMLHPSSEFFINMGRLT-----KVIKVOALPSKKRIQT 244
Db 699 KSIRDLFADML-FGKSLESVNDSPFKIKNGSFTLKAGDNINLLPNVHSLITKNVGYOI 757
Oy 245 LOSNLIARCFWFTVESGLIENHEGRKAYGAVLISSPOLG----- 286
Db 758 VAVNHIDARLITALELQNTVPSNPK-----PVIKSPVELSKSLFEVAKTIFENSVOI 810
Oy 287 ----HAFIDNRVLPLELDQIRLPFNTSTPQETLSIRHPD 324
Db 811 LKKEYTFKDNLKFFPKADGSSRLFDLPDKDQRIAPFAFVD 852

```

RESULT 38

```

T33079
Hypothetical protein F56A6.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 23-Mar-2001
C/Accession: T33079
R/Murray, J.; Kohling, T.; O'Neal, D.; Wilson, R.
Submitted to the EMBL Data Library, May 1998
A/Description: The sequence of C. elegans cosmid F56A6.
A/Reference number: Z21279
A/Accession: T33079

```

```

A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1846 <MUR>
A/Cross-references: EMBL:AF067217, PIDN:AAC17015.1; GSPDB:GN00019; CESP:F56A6.2
A/Experimental source: strain Bristol N2; clone F56A6
C/Genetics:
A/Map position: 1
A/Intons: 71/1; 165/3; 247/2; 335/2; 356/2; 369/3; 441/3; 497/3; 539/1; 607/3; 636/1; 661/1; 1714/3; 1756/2
C/Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase

```

```

Query Match 5.2%; Score 98; DB 2; Length 1846;
Best Local Similarity 23.3%; Pred. No. 22;
Matches 71; Conservative 34; Mismatches 92; Indels 108; Gaps 16;

```

```

Oy 9 DPYIKLKLKLRQSLSEFQNSLOKAVSTP-----YSYRIIQQENKKEKQALA 60
Db 217 NPKY-----ARLYFQS---KRLGSLPHIFALADVCYHNMMLIKEN----- 254
Oy 61 RHCISL-----LEFKNLLFVHLISLISNGQSGSDMAVST-----PFRNLMY 108
Db 255 --QCVVLSGESGSGKTESINLHLSLSKSGSTGCTGCTILSAGVLEAFNAVTLT 312
Oy 109 RLSSRSFLMKSYCPRFF-LDYLEAFGLSDFLDQHVAKFFLEF-----H 154
Db 313 NNNSRFG-----KFIKINREN-GMVS-----ANVEIYLLKSRITFOTGERNVH 359
Oy 155 FSYYPVSG-----FVAPHOYSLIDRYFPPLASVNRLLDKNFSLTPDLIHLGL 204
Db 360 VFYLLLEGADDEBERKKYFLKPKHDYKTLNQNEPPLALGVN-----ERNEF-----DRLR 408
Oy 205 HVFPLHPSFSEFFINMGRLFTKVIKVOALPSKKRIQTLQSNLIARCFWFTVESGL 264
Db 409 HA-----MSSVGVCAKTQOTIFGLISAV-----LLGNITYIRKHGYHSDSGY 452
Oy 265 IENHE 269
Db 453 IENHE 457

```

RESULT 39

```

S68856
Cytochrome P450 2L - spin lobster
N/Contains: oxidoreductase (EC 1.-.-.-)
C/Species: penultima atgus (spin lobster)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C/Accession: S68856; S74194
R/James, M.O.; Boyle, S.M.; Tripathi-Rosenthal, H.G.; Smith, W.C.; Greenberg, R.M.; Si Arch. Biochem. Biophys. 329, 31-38, 1996
A>Title: cDNA and protein sequence of a major form of P450, CYP2L, in the hepatopancreas.
A/Reference number: S68856; MUID:96201120; PMID:8619632
A/Accession: S68856
A/Molecule type: mRNA
A/Residues: 1-492 <JAM>
A/Cross-references: EMBL:U44826; NID:G1304739; PIDN:AAB03106.1; PID:G1304740
A/Experimental source: hepatopancreas
A/Accession: S74194
A/Molecule type: protein
A/Residues: 1-10, 'X', 12-39 <JAM>
C/Genetics:
A/Map position: 1
A/Intons: 71/1; 165/3; 247/2; 335/2; 356/2; 369/3; 441/3; 497/3; 539/1; 607/3; 636/1; 661/1; 1714/3; 1756/2
C/Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase

```

Query Match

```

5.1%; Score 97; DB 1; Length 492;
Best Local Similarity 19.9%; Pred. No. 4.6;
Matches 74; Conservative 56; Mismatches 106; Indels 136; Gaps 18;

```

```

Oy 56 KQALARKKISLIEFFKILLFVHLLSKNRCGSDMAVSTPFRRNLMVLLSSRSF 115

```

```
Db 79 KTAISKFECDRDPDYTFKLF-----GEGNDVGUV-----FSNGV----- 113
Qy 116 SLWKSYPFRFFDYLEAFGL-----LSDFLDHQAIVKFFPELETHFSY---YFVS----- 161
Db 114 -MWOTH-RRFTLRDLGMRKSRLEAAIQHEAACLVQELKKHTDQMPLPKSLINLAVLN 171
Qy 162 ---GVPAPHQVLSLQD---RYPIASVNRITLDKNFSLTDPDLHDLGHVWLL--HPS 213
Db 172 VIKVLADHRY--SLQDQGOYFTQLLTNTDMMQGFAL-----NLFNLPWLLMITPD 223
Qy 214 FSEFPIINRGRLFTKVIKQVQALSKQRITQTLQSNLIAIVRCFWFTVESGLIENHGRKA 273
Db 224 FVKNMGVRLRDGVC-----LKDYNKT-----IDNVRVLPL----- 253
Qy 274 YGAVLISPOELGHAF-----IDNVRVLPL----- 298
Db 254 -ATLDPNPKDLDAYLIDQERKEDPLSTWNIETVRAVIMDLFGAGTETSTNIRWTIL 312
Qy 299 -----ELDQIIRLPFNFTSTPQTLTSIRHFDVELTSKLEWMLDQGLLESPL----- 347
Db 313 YLMKYPEVQAKIQREIDAAVRGTLPSLEHKOKLAYFEATIHV--HRIVSLVPLGVSHY 370
Qy 348 YNQEKYLSGFVEV 359
Db 371 TNQDTLAGYRL 382

RESULT 40
G86582
exodeoxyribonuclease V, beta [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G86582
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; UID:20330349; PMID:10871362
A:Accession: G86582
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1050 <STO>
A:Cross-references: GB:BA000008; NID:98979110; PIDN:BAA98945.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: zwf
C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 97; DB 2; Length 1050;
Best Local Similarity 22.1%; Pred. No. 13;
Matches 77; Conservative 51; Mismatches 117; Indels 104; Gaps 19;

Qy 46 RILOKENKEQALAHKICISIL--BFFKNLLF---VHLLSLKQREGCSTDMA----- 95
Db 137 RLIIKHNPAHTSOLVHLHITNYLKLQWKNVLFQEQFHLLAVRYNVTSKHTSSLDKLLA 196
Qy 96 ---VSTPPFNR-----NLWY-RLLSRSFLSKSYCPFFLDYLEA-----FG 134
Db 197 SVTQPISSYFSSRVERLEQISLWHQIYNSLLEIPK---QVFLDQLTAHISGKKQKPS 252
Qy 135 LLSDFLDHQAIVKFFPELETH---FSYVPVS-GFVAPHQVLSLLODRYFIAS----- 182
Db 253 ILDDL--HHFVDDLYTSETSHSLSFFKIAETFNFKHIA-----RYKPCAFTVLNMS 305
Qy 183 -VNRITLDKNF-----SLTDPDLHDL-LGHVWLLHPSFSEFFINMGRFTKVIKQVAL 235
Db 306 WVERTFECNLDRIFTNLLVDLOEQYIKQNTYFW--LSPDESVF-----ALEKLLS 354
Qy 236 PSKKQRITQTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGH---AFIDN 292
Db 355 SBAQVQVAL-----REQYQLVLIDFQDQDKQOWISFN 389
Qy 293 VRVLPLELDQIIRLPFNFTSTPQTLTSIRHFDVELTSKLEWMLDQGL 341
Db 390 LFISPKFTGSL-----FLIGDPKQISVEWSADLPYLTAKGSFSDKQL 434
```

```
RESULT 41
F86520
glucose-6-P dehydrogenase [imported] - Chlamydothilla pneumoniae (strain J138)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: F86520
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; UID:20330349; PMID:10871362
A:Accession: F86520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:BA000008; NID:98978611; PIDN:BAA98944.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: zwf
C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 96.5; DB 2; Length 512;
Best Local Similarity 18.8%; Pred. No. 5.3;
Matches 83; Conservative 57; Mismatches 124; Indels 177; Gaps 22;

Qy 20 LRQSLSLFFNQSQSLQRAY-----STPYSYRIILOKENKEQALAHKICISILE 69
Db 113 LKDSL-----EDLQKTYGTRGNRLFVLSLTPPQVFSRIENLNKHK----- 152
Qy 70 FFKNLLFVHLLSLSKQREGCSTDMAVSTTP-----FNRLWY----- 108
Db 153 ---LFY-----KNQDQKPSVRIIEKPGRLDSAKOLOQCINENLNENSVYHIDH 201
Qy 109 -----RLLSRF-----SLWKSYPFRFFDYLE-----AFGLSDFLDHQAIVK 147
Db 202 YLGRKTYQNLTITFANTIFESCWSQ---YIDHVOISLSETIGISRGNFPEKSGMLR 257
Qy 148 -----PFELETHFSYVPVS-GFVAPH---OYLSLLODRYFI---ASVMR----- 185
Db 258 DMVQNHMMQLLCLLTLEPPTFDADIRKEKIKILO-RISFSEGSIVRGQYGPVTQVQ 316
Qy 186 -----TLKDNFSLTPDLIHDLLGHVWLLHPSFSEFFINMGRFTKVIKQVAL 235
Db 317 VSVLGYREENVDKSRVETVYVAKTVINNPRWLGVP---FYLRAGKRLAKKSTDISII 372
Qy 236 PSKKQRITQTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELGHAFIDNVR- 294
Db 373 FKKS-----PYNLFAAECSRCPIENDLL-----IIRIQPDE-GVALKFNCKV 414
Qy 295 -----VLPLELDQIIRLPFNFTSTPO-----ETLFSIRHFDLV 327
Db 415 PGTNVIRVPKMDPRYDSYFOITTPAYERLLCDCIIGDRTLFTGDEVMASWKLFTPV 474
Qy 328 ELTSKLEWMLDQGLLESIPLY 348
Db 475 E-----EW--DQDSFSPFPNY 488

RESULT 42
C72103
glucose-6-P dehydrogenase - Chlamydothilla pneumoniae (strain CWL029)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72103
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; UID:99206606; PMID:10192388
A:Accession: C72103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <ARN>
```

A:Cross-references: GB:AE001609; GB:AE001363; NID:g4376502; PID:AA018391.1; PID:g437651
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: zwf
 C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 5.1%; Score 96.5; DB 2; Length 512;
 Best Local Similarity 18.8%; Pred. No. 5.3; Mismatches 124; Indels 177; Gaps 22;
 Matches 83; Conservative 57;

QY 20 LKOSLSLFQNSQSLQRAY-----STPYSYRIILQENKEKQALAHKICISILE 69
 DB 113 LKDSL-----EDLDKTYGTRGNRLFYLLSTPPQYSRIEILNHRK----- 152

QY 70 FFKNLLFVHLISLKNQEGCSTDMAVSTPF-----FNRNLMY----- 108
 DB 153 -----LFY-----KNQDQKPMGRVIEKPFGRDIDSAKOLOCCINENLINSYHIDH 201

QY 109 -----RLSSRF-----SLKSYCPRFIDYLE-----AFGLSPDLHQAIVIK 147
 DB 202 YLGETVQNLITFRANTIFESGNSQ-----YIDHVISLSETIGIGSGKNGPFKSGMLR 257

QY 148 -----PFLETHFSYYPVSGVAPH-----QYLSLDQRYFPI-----ASVWR----- 185
 DB 258 DMVQNMHMOQLCLTMEPTTFDDELREKIKILO-RISPFSEGSISVROQYGRVYQG 316

QY 186 -----TLDKNBSLTPTDLIHDLGHVPLHPSEFPINMGRFLTKVIEKQAL 235
 DB 317 VSVLGYREBNVDKDSRVEYIVALKYINPRMLGVF-----FYLRAGKRLAKSTDISI 372

QY 236 PSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYAVLISPOELGHAFIDNVR 294
 DB 373 FKKS-----PYNLFAEECSRCPINDLL-----IIRIQPE-GVALKFNCKV 414

QY 295 -----VPLELDQIIRLPENTSTPO-----ETLFSIRHPDELV 327
 DB 415 PGTNNIVRPVKMDFRYSYFQTTPEAYERLLCCLIGDRTLFTSGDEVMAWKLFPVL 474

QY 328 ELTSKLEWMLDQGLLESIPLY 348
 DB 475 E-----EW-----DQSSSPSPFN 488

RESULT 43
 562468
 probable membrane transporter - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
 C:Accession: J138579, S62468
 R:Badcock, K.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, October 1995
 A:Reference number: Z11745
 A:Accession: J138579
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-529 <BA2>
 A:Cross-references: EMBL:Z54354; NID:g1019398; PID:CAA91178.1; PID:g1019411; GSPDB:GN00
 A:Experimental source: strain 972h-; cosmid c2G11
 C:Genetics:
 A:Gene: SPAC2G11.13
 A:Map position: 1L

Query Match 5.1%; Score 96; DB 2; Length 529;
 Best Local Similarity 22.0%; Pred. No. 6.1; Mismatches 127; Indels 72; Gaps 11;
 Matches 66; Conservative 35;

QY 120 SYCPRFPIVLEAFGLSD--FLDHQAVIKFLETHFSYYPVSGVAPHGYLSLDQRY 177
 DB 133 SWSPSFLLKIVPLVLDNFISSQCYDSF-LPIFLAFYITR--GPITLSALQDER 189

QY 178 FFIASVMTLKDKNFSLTPLDI-HDLIGH-----VPMPLHPSSEFP-- 218
 DB 190 DDUDSYITNTTIDSEEPYLLHSLTLNESAPPAVDPEBHKAKTAARLSSVGFOSFPGA 249

QY 219 -IMNGRLFTKVIEXQALPSKKORIOTLOSNIATVRCFWFTVESGL-----IENHEG 270
 DB 250 ALFQIIFPIILKTNND-----ILPFTVVCSCMLISTPLCTIVLPPVENSS 302

QY 271 RKAYGAVLISPOELGHAFIDNVRVPLELDQIIRLPN-----TSPQETLS----- 319
 DB 303 -DALITLVNSVESYHSFKAMSSISIRLFLERFLINCIGTSSSAVIGKARLNS 361

QY 320 -----IRHDELVELTSKLEWMLDQGLLESIPLYNGEKLISGE 358
 DB 362 NFQITLLGNGISFALLGTVIIPLYTEVFNLSIQVVMIIISILLPMAPLYGLGYIPGE 421

RESULT 44
 H72041
 exodeoxyribonuclease V, beta - Chlamydomonas pneumoniae (strain CWL029)
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C:Accession: H72041
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammell, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: H72041
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1050 <ARN>
 A:Cross-references: GB:AE001655; GB:AE001363; NID:g4377039; PID:AA018877.1; PID:g43
 A:Experimental source: strain CWL029
 C:Genetics:
 A:Gene: recB

Query Match 5.1%; Score 96; DB 2; Length 1050;
 Best Local Similarity 22.1%; Pred. No. 15; Mismatches 117; Indels 104; Gaps 19;
 Matches 77; Conservative 51;

QY 46 RIILOKENKEKQALAHKICISIL--EFPKNLF--VHLISLKNQEGCSTDMA----- 95
 DB 137 RLHKNPALTHSOLVLIHNTVNLKODLWKVLPQEPHLLAVRYNTSKTSLVDKLLA 196

QY 96 -----VSTPEFNR-----NLMY-RLSSRFSLMKSYCPRFIDYLEA-----FG 134
 DB 197 SYTQPISSYSSRVERLEQISLMHQIYNLSLEIRK-----QVPLDQLTAHISGFKKQPS 252

QY 135 LSGPDLHQAIVIKFLETH--PSYYPVS-GFVAPRQYLSLDQRYFPIAS----- 182
 DB 253 ILDDL-HHVDLVYSEHSSLSFPAKIAATPFKIRLA-----RYKCAAFVLENN 305

QY 183 -VNRITDKDNF-----SLTPDLIDL-LGHVPMPLHPSSEFPINMGRFLTKVIEKQAL 235
 DB 306 WERTLSEFCULDRIFNTLVLDQETLKONYTPW-LSDDESVP-----ALEKLSS 354

QY 236 PSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYAVLISPOELGH--AFIDN 292
 DB 355 SEAPVVOAL-----RQYOLVLIIDEFOPTDKQWSTFSN 389

QY 293 VRVLPLELDQIIRLPENTSTPOETLFSIRHDELVELTSKLEWMLDQGL 341
 DB 390 LFLSPKFTGSL-----FLIGDPKOSIYEWRSADLPYLTAKSSFSBDKOL 434

RESULT 45
 T41933
 large tegument protein - human herpesvirus 7 (strain J1)
 C:Species: human herpesvirus 7
 A:Variety: strain J1
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T41933
 R:Nicholas, J.
 submitted to the EMBL Data Library, December 1995
 A:Description: Determination and analysis of the complete nucleotide sequence of hum.
 A:Reference number: Z22022

A:Accession: T41933
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2059 <N1C>
A:Cross-references: EMBL:U43400; PIDN:AAC54693.1
A:Experimental source: strain J1
C:Genetics:
A:Note: U31
C:Superfamily: varicella-zoster virus 240K tegument protein

Query Match 5.1%; Score 96; DB 2; Length 2059;
Best Local Similarity 20.6%; Pred. No. 37;
Matches 90; Conservative 62; Mismatches 164; Indels 120; Gaps 23;

QY 13 ILKIALKRLQSLFFQNSOS-LQRAYSTPYSYVYRIILQENKEKQALARKK--CISILE 69
DB 1265 ILKWLIVPVKELNTFFVATMSERGEVVPDYKHFR-ALEYINSKYIEIKNICNEIIE 1323
QY 70 FPKNL-LFVHLLSLSKNQREGCSTD----MAVSTPFNNLWYRLLSRF--SLWKSX 121
DB 1324 NTDNTEKSLTIKQIDPNRTAGKQKQFDYLSKIULTASTNQO-----QRYKEQKKQY 1377
QY 122 CRRFLDYLEAFGLSLDFLDHQAIV-----KFELETHFSYYP-----VSGF 163
DB 1378 FD-LLDNIAHFRFAPDFNQNLILKLDKDKFKTLRTDTVERPNDLDDTFVSSMNVNF 1435
QY 164 VAPHQYLS-----LLQDRYFPI-ASVMTLTKDNFSL-----TP- 196
DB 1436 LQALEALSHFVQAQNFQNLVTEQADLPQTNFIPVELSTVKTIPKSDINLRMKIHTPO 1495
QY 197 -----DLIHLGLHVPWLLHPSSEFFINMGRL-FTKVI--EKVOALPSKKQRI 242
DB 1496 TPFQVDSVFNTQLIVDEKG-IPQVFNVFNHIVKFFALNVKKIIPDKVLNVSTKYI 1554
QY 243 QFQSLNLAIVRCFW-FTVBSGLIENHGR-----KAYGAVLIS-----SPQ 283
DB 1555 LTTLSILSVKSPKKEIINFDLTSYFQCKAEFTFQNVFPIINLKIYIITQAMSVTSD 1614
QY 284 ELGHAFIDNVVPLEDQIIRLPFTNTSPQSTLSIRHPDELVELSKLEWMLDOGLLE 343
DB 1615 EQHSHF-----ELPKEKFSLLIANN-----PEFLF-----GSLQCPVDLAINS 1653
QY 344 SIPLYNQSKYLSGFEV 359
DB 1654 LIPLLEKKKXYTFTI 1669

RESULT 46
E70066
hypothetical protein ywq8 - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: E70066
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabbet, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, C.V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maues
Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, E.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, R.; Segiguchi, J.; Sekowska, A.; Seroh
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,
I.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E70066
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-536 <KUN>

A:Cross-references: GB:Z99122; GB:AL009126; NID:G2636029; PIDN:CAB15644.1; PID:G2636152
A:Experimental source: strain 168
C:Genetics:
A:Gene: ywq8
C:Superfamily: Bacillus subtilis hypothetical protein ywq8

Query Match 5.1%; Score 95.5; DB 2; Length 536;
Best Local Similarity 21.8%; Pred. No. 6.8;
Matches 57; Conservative 30; Mismatches 93; Indels 81; Gaps 12;

QY 107 WYRLLSGRFSLWKSYP-----PFLLDYLEAF-----GLSDPLDHOAV---IKF 148
DB 142 WLAFDSEFSLWQARTPEGSQNMQGLYGYLSALKKHAPKPELKSLYQIHSIAVCLRM 201
QY 149 FELETHFSYYPVSGFVAPHQYLSLLQDRYFPFIASVMT-----LQDNFSLTPDL 198
DB 202 FTLEAGKLNPEQDFSLNYPVQQLMDTIYSSIDKLTIALYSPALDFLDK-----TPDV 256
QY 199 IHDLLGHVPWLLHPSSEFFINMGRLFTKVIKVOALPSKKQRIOTLOSNIILAIIVRCFW 258
DB 257 IRLHL-----LKEIFQYERI-----RVFGEISALLSRPK-----WV 289
QY 259 TVESGLIENHGRKAYGAVLISPOELGHAFID---NVRVPLELQDIIRLPFTNTSTPQ 314
DB 290 AREQELILKKEAGRR-----FSPLEQFGLHLEFLQKNDVIFBEADQF-----PP 334
QY 315 ETLFSIRHFDLVELTSKLEW 335
DB 335 EAL--PYTFQWLSEMTAKKDW 353

RESULT 47
B84846
probable zinc proteinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: B84846
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bonito, M.I.; Town, C.D.; Fujii, C.Y.; A
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: B84420; MUID:20083487; PMID:10617197
A:Accession: B84846
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-970 <STO>
A:Cross-references: GB:AE002093; NID:G2335108; PIDN:AAC02769.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g41790
A:Map position: 2
C:Superfamily: inaeulysin

Query Match 5.0%; Score 95; DB 2; Length 970;
Best Local Similarity 21.7%; Pred. No. 17;
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

QY 9 DPKYILKALKRQSLSPFQNSOSLQRA-----YSTPYSYVYRIILQENKEKQALAR 61
DB 421 NPALVQKVVDLSNPNFIPWESQKEGQTDKAEFWYNTAISLEKITSS-----469
QY 62 HKCISILEFFKNLLFVHL-----LSLSKQREGCSTDMAVSTPFNNLWYR 109
DB 470 ----TIQSWQASPDVHLHLPAPNVIPDLSL-KDADDKETVPVLURKTFP---SRIMYK 522
QY 110 LLSRFSPLWKSX-----CP-----RFFLDYLEAF-----133
DB 523 -PDMFSKPKAYVMDPNCPLAVSSPDAAVLDTIFRLMDYLNLEYAYYAQVAGLYGVS 581
QY 134 -----GLLSDF---LDHQAVIKFFFELETHFSYYPVSGFVA 165
DB 582 LSDNGFELTLGYNHKLRIILLETVVGVKIANFENKPDRAVIK---ETVTKEYQNYKFRQ 637

QY 166 P-HO---YLST-LDRYFPIASVWRTLDKDNFSLTPDLIHDLGH-----VPMILH 211
 DB 638 PTHQMYVCSLIDQTP-----WTEL--DVLSHLEADVAKVPMILS 681
 QY 212 PPSFSPFI-----NMGRLFTKVIKY-----QALPSKQRIQTLOSNIATVRC 255
 DB 682 RTFECYTAGVNNENAEASVWKHEVLENDPKPCRPFPSS-----QHLTRVVKL--- 733
 QY 256 FWFVTEGSLIENHEGRKAYGAVLISSFQELGHAFLDNVAVLPLELDQIIRLP-FNTSTPQ 314
 DB 734 -----GEGMKYFPHQDGSNPSDENSALVHYIQVHRDPSMNIKQLPGLGLVAKQ 781
 QY 315 ETLPSIRHDELVELVTSKLEWMLDOGL 341
 DB 782 ATPHQLRTVEOLGYITA-LAQRNDSGI 807

RESULT 48

S8B8 protein - Yeast (Saccharomyces cerevisiae)

N.Alternate names: cyclin-dependent kinase 5 activator homolog; probable membrane protein

C.Species: Saccharomyces cerevisiae

C.Date: 19-Mar-1997 #sequence revision 25-Apr-1997 #text change 20-Jun-2000

C.Accession: S74293, S19496, S19495, S19766; C57062; S53594; S55868; S68611; S59988

R.Wedder, H.; Wandute, R. #sequence revision 25-Apr-1997 #text change 20-Jun-2000

Submitted to the Protein Sequence Database, September 1996

A.Reference number: S74288

A.Accession: S74293

A.Molecule type: DNA

A.Residues: 1-1427 <WED>

A.Cross-references: EMBL:X59720; NID:G1907116; PID:G1907218; MIPS:YCR081w

R.Ballesta, J. G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.

Submitted to the Protein Sequence Database, March 1992

A.Reference number: S19486

A.Accession: S19496

A.Molecule type: DNA

A.Residues: 202-863 <BAL>

A.Cross-references: EMBL:X59720; MIPS:YCR080w

A.Accession: S19495

A.Molecule type: DNA

A.Residues: 1-1427 <FEL>

A.Cross-references: EMBL:X59720; MIPS:YCR080w

A.Note: This sequence has been revised in reference S55867

A.Note: YCR080w

R.Feldmann, H.; Mannhaupt, G.; Vetter, I.

Submitted to the Protein Sequence Database, March 1992

A.Reference number: S19429

A.Accession: S19766

A.Molecule type: DNA

A.Residues: 531-1427 <FEL>

A.Cross-references: EMBL:X59720; MIPS:YCR080w

R.Hengartner, C.U.; Thompson, C.M.; Zhang, J.; Chao, D.M.; Liao, S.M.; Koleske, A.J.; Ok

Genes Dev. 9, 897-910, 1995

A>Title: Association of an activator with an RNA polymerase II holoenzyme.

A.Reference number: A57062; MUID:95293223; PMID:7774808

A.Accession: C57062

A.Status: not compared with conceptual translation

A.Molecule type: DNA

A.Residues: 202-1427 <HEN>

R.Jimenez, A.

Submitted to the EMBL Data Library, December 1992

A.Reference number: S53589

A.Accession: S53594

A.Molecule type: DNA

A.Residues: 1-133, 'RFV', 136-137, 'FDMS', 142, 'LASLYLTA' <JIM>

A.Cross-references: EMBL:X59720

R.Rodriguez-Cousino, N.; Lill, R.; Neupert, W.; Court, D.A.

Yeast 11, 581-585, 1995

A>Title: Identification and initial characterization of the cytosolic protein Ycr77p.

A.Reference number: S55867; MUID:95373282; PMID:7643349

A.Accession: S55868

A.Molecule type: DNA

A.Residues: 1-133, 'RFV', 136-137, 'FDMS', 142, 'LASLYLTA' <ROD>

A.Cross-references: EMBL:X59720
 R.Huang, Q.O.; Lee, K.Y.; Wang, J.H.
 FEBS Lett. 378, 48-50, 1996
 A>Title: A novel yeast protein showing specific association with the cyclin-dependent
 A.Reference number: S68611; MUID:96140706; PMID:8549800
 A.Accession: S68611
 A.Molecule type: protein
 A.Residues: 701-1011, 1013-1056 <HUA>
 C/Genetics:

A/Gene: SCD:SRB8

A/Cross-references: SCD:S0000677; MIPS:YCR080w; MIPS:YCR081w

A/Map position: 3R

C/Superfamily: Saccharomyces cerevisiae SRB8 protein

C/Keywords: nucleus; transmembrane protein

F/277-293/Domain: transmembrane #status predicted <TM1>

F/353-369/Domain: transmembrane #status predicted <TM2>

F/566-582/Domain: transmembrane #status predicted <TM3>

F/693-709/Domain: transmembrane #status predicted <TM4>

F/724-740/Domain: transmembrane #status predicted <TM5>

F/1002-1018/Domain: transmembrane #status predicted <TM6>

F/1081-1097/Domain: transmembrane #status predicted <TM7>

F/1169-1185/Domain: transmembrane #status predicted <TM8>

Query Match 5.0% Score 95; DB 2; Length 1427;

Best Local Similarity 18.8% Pred. No. 28;

Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YLTKIALKRLQSLSLFPQNSGLRAYSVPYSYRII-----LQENKEKQALAHKCI 65
 DB 214 YLLE---KLFDNTNHYNSQOL-RTWKROLSYFKLGNCSYLRINBE---IFHMLV 266
 QY 66 SLIEFFKNILF-----VHLISLKNQREGCSTDMAVST-----PFF----- 102
 DB 267 EFINKMENFEPLSLHILMIFWMDICQIDNAPVAATITYSQKEPFLVTKITDMLHK 326
 QY 103 -----NNLWYRLSSRFL-----WKS 120
 DB 327 YVIVSSKSMINDENYIINDIKNNKIKLITSLIKIFQEOSLEVFIFPTSNWEI 386
 QY 121 YCRPFF-----LDYLEAFGLSDFLDHOAVIKFPELTHRSYVPGFVAP 166
 DB 387 KYPLLEFIVSNADPNQNSDMKKLELISYRNBSLKNSSIR-----NIVMSASMAN 437
 QY 167 HGYLSILDRYFPIASV-----RTLDKXN-----FSLTPDLIHDLGHVP 207
 DB 438 DFLITVTCQFPKLSQIOLNCIDQFTKLD-DNPEFPMPTVYVQNFVTHKIIQLL 496
 QY 208 WLHPG--PSEFIMNGRLFTVIEKVQALPSKQRIOTLOSNI--IAVRCFPTVBSG 263
 DB 497 WSHPSKQPDHYSN-----QVAKLIL-----KINSDEDLHEFOEDAIWSLVFO 544
 QY 264 LIENHEGRK-----AYGAV-----LIISP-----QELGHAFLD-- 291
 DB 545 LAKNFSQKQKVSYYMPSLYRLNLITVYGIKVTYIRKLSSGLVYQDSNDKRVHQ 604
 QY 292 ---NVRVLPLELDQIIRLPFNTSTPQETLPSIRHDELVELTSKL----- 333
 DB 605 LLINLKSLPKMSQYNNVLRNMEYDKYEIFPFDDLVITFOIKRILINDITNLOS 664
 QY 334 -----EWMLD--QGLESIPLYNOEKYLSGFEVLC 361
 DB 665 KTLPSIKIMVAEWYLSHLCSGILSSV--NRTVLKIKFIKFC 703

RESULT 49

A60026

cell communication-mediating membrane protein ptra2a - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 03-Dec-1993 #sequence revision 18-Nov-1994 #text change 08-Oct-1999

A/Accession: A60026

R.Kuwahara, P.E.; Okuma, P.G.; Kimble, J.

Mol. Biol. Cell 3, 461-473, 1992

A>Title: tra-2 encodes a membrane protein and may mediate cell communication in the

A:Reference number: A60026; MUID:92360913; PMID:1498366

A:Accession: A60026

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1475 <KUM>

A:Cross-references: GB:S42187; NID:9253436; PIDN:AAB2845.1; PID:9253437

A>Note: sequence extracted from NCBI backbone (NCBIN:110873, NCBI:P110874)

C:Keywords: transmembrane protein

Query Match 5.0%; Score 95; DB 2; Length 1475;

Best Local Similarity 16.7%; Pred. No. 29;

Matches 75; Conservative 65; Mismatches 120; Indels 188; Gaps 19;

QY 1 VHYCERTLDPKYILKIALKLRQ-----LSLFFQNSQSLQRAYSTPYRYII 48

DB 77 LHTQHGVLNNISNLKIKPRQTYTLNDVCFKPHITIFQSSSDQNEY--PHYIQRLL 134

QY 49 LQENKEQALARKHCISILEFFKNLLFVH--LLSLSKNQREGCSTDMVAVVSTPFFNRNL 106

DB 135 L-----ENQRLSPCLIVTPLNCFYDIYRIGHSINWKN-----TDFLNR-- 174

QY 107 MYRLSSRFLSKSYCPFPFLDYLEAFG-----LLSDFLDHQAVIKFFEL 151

DB 175 --RLRNS-----YIEAIGENDERPYVKSNGPSLIKSWADH-----MFDL 212

QY 152 ETHFSYYPVSGF-----VAPHQY-----LSLQDRYFPPIASV 183

DB 213 -----PSKFTNSTKDALFOKIKLWLLSIEPRQKTCASIIHSCDTPLDSEHYFNICTD 265

QY 184 MRLDK-----DNPSLTPDLIHLGHVPWLLHPSFSEFFINMGRLETKVI 229

DB 266 MOSVDNFAEKKTEKLEDDVEEPAMLCDDVQOEFIEMQOE-----LEIKMYSHVT 318

QY 230 EK-----VQALSKKQRIQTLQSNLIAV----- 253

DB 319 EKPDPYVNVNQCCKPHDLNLTGIEFFPGSFSSTKSDQFTMTQTEIVLLTPEMLLSAM 378

QY 254 -----RCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDN-----VRVPLE 299

DB 379 QHSDPVGFIW--TIEKABELIHEFLAL-----KEETKFKENRMSKMRVTSRV 429

QY 300 LDQIIRLPNTSTPQETLFSIRHDELV 327

DB 430 LDNTVTTKLOSFSEKQTIHFVNVVHSLI 457

RESULT 50

C81321

Probable cation-transporting ATPase Cj1161c [imported] - Campylobacter jejuni (strain NC

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: C81321

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: C81321

A:Status: preliminary

A:Molecule type: DNA

A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAB73415.1; PID:9696859

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj1161c

C:Superfamily: Thermotoga maritima P-type cation-transporting ATPase; ATPase nucleotide-

F:8-37/Domain: heavy-metal-associated homology <HMA>

F:519-657/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 5.0%; Score 94.5; DB 2; Length 699;

Best Local Similarity 22.5%; Pred. No. 12;

Matches 76; Conservative 54; Mismatches 101; Indels 107; Gaps 18;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 12:09:28 ; Search time 25 Seconds
(without alignments)

600.577 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCERTLDPKYLKIALKL.....ESTPLNQEKLSGPEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	99.8	362	1	AAAH_CHLPN
2	236.5	13.6	289	1	PHAH_VISCH
3	235.5	12.5	262	1	PHAH_PSEAE
4	231	12.2	445	1	TRSH_CHICK
5	230.5	12.2	444	1	TRSH_HUMAN
6	229.5	12.1	491	1	TY3H_PHASP
7	227.5	12.0	444	1	TRSH_RAT
8	226.5	12.0	447	1	TRSH_MOUSE
9	225.5	11.9	444	1	TRSH_RABIT
10	225.5	11.9	488	1	TY3H_ANGAN
11	225.5	11.9	498	1	TY3H_MOUSE
12	223.5	11.8	481	1	TRSH_XENIA
13	223	11.8	498	1	TY3H_RAT
14	221.5	11.7	490	1	TY3H_BOVIN
15	221	11.7	275	1	PHAH_RHILO
16	220	11.6	528	1	TY3H_HUMAN
17	219	11.6	508	1	TY3H_DROME
18	216.5	11.5	465	1	TY3H_SCHMA
19	214	11.3	453	1	PHAH_RAT
20	213.5	11.3	452	1	PHAH_HUMAN
21	212.5	11.2	453	1	PHAH_MOUSE
22	201	10.6	313	1	PHAH_RALSO
23	196.5	10.4	297	1	PHAH_CHRVO
24	194	10.3	452	1	PHAH_DROME
25	189.5	10.0	294	1	PHAH_CAUCR
26	178.5	9.4	524	1	TY3H_CAEEL
27	175.5	9.3	457	1	PHAH_CAEEL
28	106.5	5.6	1501	1	NINC_DROME
29	104.5	5.5	2331	1	RRPL_MABVP
30	102	5.4	563	1	YAS8_SCHPO
31	100.5	5.3	1157	1	N133_YEAST
32	100.5	5.3	2212	1	RRPL_EBOMZ
33	98.5	5.2	1024	1	Y075_MYCGE

34	97	5.1	492	1	CPL1_PANAR
35	97	5.1	1050	1	EX5B_CHLPN
36	96.5	5.1	512	1	G6PD_CHLPN
37	96	5.1	529	1	YABD_SCHPO
38	95.5	5.1	2059	1	TEGQ_HSV7J
39	95.5	5.1	1398	1	TOP2_PLAFK
40	95	5.0	1427	1	SRB8_YEAST
41	95	5.0	1475	1	TRA2_CABEL
42	93.5	4.9	650	1	Y096_MYCGE
43	92.5	4.9	466	1	Y065_MYCGE
44	92	4.9	447	1	FLP_KIULA
45	92	4.9	526	1	MATK_ARATH
46	92	4.9	544	1	MATK_MAIZE
47	92	4.9	4563	1	APB_HUMAN
48	91.5	4.8	293	1	SPEE_METJA
49	91	4.8	2183	1	RRPL_MEASA
50	91	4.8	2183	1	RRPL_MEASE
51	90.5	4.8	114	1	ISP_BUCAI
52	90.5	4.8	616	1	ORC3_YEAST
53	90.5	4.8	1102	1	TR13_YEAST
54	90.5	4.8	2331	1	RRPL_MABVM
55	90	4.8	835	1	VP3_ROTSL
56	90	4.8	863	1	APCE_GALSU
57	90	4.8	892	1	YLB3_CAEEL
58	90	4.8	4590	1	FATH_HUMAN
59	89.5	4.7	2136	1	YCF2_MARPO
60	89	4.7	1037	1	YQJ8_YEAST
61	89	4.7	1197	1	EVGS_ECOLI
62	89	4.7	1197	1	EVGS_ECOLI
63	88.5	4.7	519	1	ARH5_HUMAN
64	88.5	4.7	891	1	ARH5_HELPY
65	88.5	4.7	1967	1	YG50_YEAST
66	88	4.7	439	1	Y204_METJA
67	88	4.7	659	1	CH4A_HUMAN
68	88	4.7	873	1	YAJ4_YEAST
69	88	4.7	2291	1	R3PB_BEV
70	87.5	4.6	1319	1	CLT1_RAT
71	87.5	4.6	454	1	WXYE_YERPE
72	87.5	4.6	542	1	MATK_ORISA
73	87.5	4.6	594	1	PEK_CANJE
74	87	4.6	324	1	CBH_LACEL
75	86.5	4.6	495	1	YNS3_ROTSP
76	86.5	4.6	986	1	EP1B_STARP
77	86.5	4.6	1202	1	RPW2_YEAST
78	86.5	4.6	1492	1	AT7A_RAT
79	86	4.6	523	1	RPN3_YEAST
80	86	4.6	1405	1	DPOA_SCHPO
81	86	4.6	1845	1	YM68_CAEEL
82	85.5	4.5	839	1	TLR4_HUMAN
83	85.5	4.5	839	1	TLR4_PANPA
84	85.5	4.5	2039	1	CCH1_YEAST
85	85	4.5	238	1	Y106_METJA
86	85	4.5	462	1	SYTM_YEAST
87	85	4.5	493	1	Y100_STRPN
88	85	4.5	1235	1	RNT1_ARATH
89	84.5	4.5	585	1	YHQ9_YEAST
90	84.5	4.5	719	1	YB62_CABEL
91	84.5	4.5	1070	1	EX5C_BUCAI
92	84.5	4.5	1748	1	POLR_ELV
93	84.5	4.5	3668	1	POLG_PEMVC
94	84	4.4	240	1	YC67_AQUAE
95	84	4.4	501	1	PIP_LACIA
96	83.5	4.4	423	1	ARGD_YEAST
97	83.5	4.4	450	1	WZYE_ECO57
98	83.5	4.4	450	1	WZYE_ECOLI
99	83.5	4.4	614	1	SGAC_RABIT
100	83.5	4.4	835	1	UBPQ_MOUSE
101	83	4.4	396	1	RT09_HUMAN
102	83	4.4	610	1	ABCI_SCHPO
103	83	4.4	652	1	MX1_RAT
104	83	4.4	764	1	DPO3_SULSO
105	83	4.4	962	1	YBX7_SCHPO
106	83	4.4	969	1	YEAC_SCHPO

Q27712	panulirus a
Q927g7	chlamydia p
Q928u6	chlamydia p
Q09812	schizosacch
P52362	human herpe
P41001	plasmodium
P25648	saccharomyc
P34709	caenorhabdi
P47342	mycoplasma
P47311	mycoplasma
P13783	kluyveromyc
P56784	arabidopsis
P48190	zea mays (m
P04114	homo sapien
Q57761	methanococc
P35975	measles vir
P12576	measles vir
P57247	buchnera ap
P54790	saccharomyc
Q03660	saccharomyc
P31352	matburg vir
P15736	simian 11 r
P35911	galdieria s
P46578	caenorhabdi
Q14517	homo sapien
P09975	marchantia
Q12496	saccharomyc
P58402	escherichia
P30855	escherichia
Q12774	homo sapien
P56105	helicobacte
P53327	saccharomyc
Q06260	methanococc
Q13619	homo sapien
P40309	saccharomyc
P18458	berne virus
Q92428	rattus norv
Q82451	yersinia pe
P12175	oryza sativ
Q9pmu0	campylobact
Q06115	lactobacilli
P35425	simian 11 r
P30195	staphylococ
Q02773	saccharomyc
P70705	rattus norv
P40016	saccharomyc
P28040	schizosacch
P34529	caenorhabdi
O00206	homo sapien
Q9ttu0	pan paniscu
P50077	saccharomyc
Q57570	methanococc
P07236	saccharomyc
Q97p44	streptococc
Q9fjro	arabidopsis
P38818	saccharomyc
Q09437	caenorhabdi
P57528	buchnera ap
P35928	erysinum la
Q01500	p genome po
Q67305	aquifex aeo
P49022	lactococcus
P18544	saccharomyc
Q8xaq5	escherichia
P27835	escherichia
P49055	oryctolagus
Q99mx1	mus musculus
P82933	homo sapien
Q92338	schizosacch
P18588	rattus norv
P59579	fulfolobus
Q10201	schizosacch
Q14077	schizosacch

107 83 4.4 1014 1 CINI_YEAST P40987 saccharomyc
108 83 4.4 1254 1 DROC_YEAST P15801 saccharomyc
109 82.5 4.4 455 1 ORCS_SCHPO O41114 schizosacch
110 82.5 4.4 542 1 MATK_HEAT P58271 triticum ae
111 82.5 4.4 545 1 YB90_YEAST P38346 saccharomyc
112 82.5 4.4 1001 1 MPD_HELPI O91373 heliobacte
113 82.5 4.4 1491 1 AT7A_MOUSE O64130 mus musculu
114 82.5 4.4 2210 1 NRPL_EBOSM Q68802 ebola virus
115 82 4.3 365 1 MLI2_LACLC P50179 lactococcus
116 82 4.3 364 1 RABG_BACSU O32299 bacillus su
117 82 4.3 367 1 PPSX_CHICK P08836 gallus gall
118 82 4.3 475 1 ATXX_RAT O9er24 ratius norv
119 82 4.3 641 1 MATK_HORVU P17158 hordeum vul
120 82 4.3 775 1 YPVA_BACSU P50831 bacillus su
121 82 4.3 819 1 UL87_HSVJ7 P52363 human herpes
122 82 4.3 1104 1 YCXA_CHLVU P56370 chlorella v
123 82 4.3 1458 1 CARB_ANASP O8YQ12 anabeena sp
124 82 4.3 1738 1 CUTA_SCHPO O9YU72 schizosacch
125 82 4.3 618 1 YCF1_EPIVI O00383 epifagus vl
126 81.5 4.3 915 1 ORC2_DROME O24168 drosophila
127 81.5 4.3 2294 1 KCQ4_YEAST P25389 saccharomyc
128 81.5 4.3 3056 1 YCF2_ARATH P56786 arabidopsis
129 81.5 4.3 3276 1 ATM_HUMAN O13315 homo sapien
130 81 4.3 298 1 SPEE_BACSU P70998 bacillus su
131 81 4.3 323 1 ERA_XYLP4 O9pb97 xyloella fae
132 81 4.3 428 1 UMEJ_YEAST P47821 saccharomyc
133 81 4.3 475 1 RUMA_HAEIN P46643 haemophilus
134 81 4.3 506 1 ATXX_MOUSE P26659 mus muscula
135 81 4.3 506 1 SYE_CHLIR O84451 chlamydia c
136 81 4.3 880 1 YB14_YEAST P33961 saccharomyc
137 81 4.3 1016 1 YMD5_CABEL P34459 caenorhabd
138 81 4.3 3830 1 FDOG_ECOLI P32176 escherichia
139 81 4.3 346 1 SACS_MOUSE O9J1C8 mus musculu
140 80.5 4.3 377 1 TRAJ_LESBO O48514 leptospira
141 80.5 4.3 377 1 PPTA_RAT O04631 ratius norv
142 80.5 4.3 787 1 ECCE_HUMAN O60344 homo sapien
143 80.5 4.3 826 1 TLR4_PAPAN O9SP22 papio anubi
144 80.5 4.3 959 1 VP2_BTIV3V Q06998 bluetongue
145 80.5 4.3 1234 1 YNX5_CABEL P34578 caenorhabd
146 80.5 4.3 1365 1 KRES_YEAST P22023 saccharomyc
147 80.5 4.3 2493 1 YBA4_YEAST P35194 saccharomyc
148 80 4.2 253 1 VG23_HSVSA O01006 herpesvirus
149 80 4.2 325 1 PK1_NPVAC P41415 autographa
150 80 4.2 325 1 YAB6_YEAST P31379 saccharomyc

ALIGNMENTS

RESULT 1
ID AAAH_CHLPN STANDARD; PRT; 362 AA.
AC 0926L3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Probable aromatic amino acid hydroxylase (EC 1.14.16.-).
GN CPN1046 OR CP0806 OR CPJ1046.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxId=83556;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CM1029;
RA MEDLINE=9920606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RT Olinger L., Gilmord J., Davis R.W., Stephens R.S.,
RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RN Nat. Genet. 21:385-389(1999).
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RA MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA Linhe O., Hickey E.K., Peterson J., Uetreback T., Berry K., Bass S.,
RA Whitte K., Weidman J., Khouri H., Craven C., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=2033049; PubMed=10871362;
RA Shiba T., Ishii K., Hattori M., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CM1029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1 - COFACTOR: Binds 1 ferrous ion (By similarity).
CC -1 - SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001685; AD19183.1; -;
DR EMBL; AE002240; AA073705.1; -;
DR EMBL; AP002548; BAA99253.1; -;
DR TIGR; CP0806; -;
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; bioplerin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR TIGRPFAMs; TIGR01267; Pheahydrox mono; 1.
DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Iron; Complete proteome.
FT METAL 200
FT METAL 205
FT IRON (POTENTIAL).
SQ SEQUENCE 362 AA; 42513 MW; 01B89B84B4FE53B CRC64;
Query March 99.8%; Score 1886; DB 1; Length 362;
Best local similarity 99.7%; Pred. No. 2e-141; 0; Indels 0; Gaps 0;
Matches 361; Conservative 1; Mismatches

OY 1 VHYCERTDPKYILKIALKRLQSLSPFONOSLOPASYSPYVYRIILQKXKROALA 60
DB 1 WHICERTLDPKYLKIALKRLQSLSPFONOSLOPASYSPYVYRIILQKXKROALA 60
OY 61 RHKICSTLEFPKULFVHLISLKNQREGCTDAVYSTPPNNMLYRLLSRFLWMS 120
DB 61 RHKICSTLEFPKULFVHLISLKNQREGCTDAVYSTPPNNMLYRLLSRFLWMS 120
OY 121 YCPFFFLDYLEAFGLSDFLDHQAIVKPELETHPSYVYVPGFAVPHQYVLSLQDRYPI 180
DB 121 YCPFFFLDYLEAFGLSDFLDHQAIVKPELETHPSYVYVPGFAVPHQYVLSLQDRYPI 180
OY 181 ASVVRITLQKDFSLTPDLIHDLGIVPWLPLHPSSEFFIMNGRLFTVIEKVALSKSKQ 240
DB 181 ASVVRITLQKDFSLTPDLIHDLGIVPWLPLHPSSEFFIMNGRLFTVIEKVALSKSKQ 240
OY 241 RIOTLQSNLAIYRCFWFVESGLLENHGGKAGAVVLSPOELGAFIDNVVPLPL 300
DB 241 RIOTLQSNLAIYRCFWFVESGLLENHGGKAGAVVLSPOELGAFIDNVVPLPL 300
OY 301 DQIRLPLFNSTQETLSIRHPDELIVLSTKEMMDQGLSSIPLYNOEKLSGFVL 360
DB 301 DQIRLPLFNSTQETLSIRHPDELIVLSTKEMMDQGLSSIPLYNOEKLSGFVL 360
OY 361 CQ 362

Db 361 CQ 362

RESULT 2

PH4H_VIBCH STANDARD; PRT; 289 AA.

AC Q9KLB8;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-monooxygenase).

GN PH4H OR VCA0828.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN NCBI_TaxID=666;

RP SEQUENCE FROM N.A.

RX STRAIN=El Tor N16961 / Serotype O1;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., McInerney M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M.;

RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.";

RT Nature 406:477-483 (2000).

RL NCBI_Catalytic Activity: L-phenylalanine + tetrahydrobiopterin + O(2) = L-tyrosine + dihydrobiopterin + H(2)O.

CC -!- COFACTOR: Binds 1 ferrous ion (By similarity).

CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.

CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AE004410; AAF96726.1; --

DR HSSP; P04177; 1TOH.

DR TIGR; VCA0828; --

DR InterPro; IPR001273; Aaa_hydroxylase.

DR Pfam; PF00351; bioprotein_H_1.

DR PRINTS; PR00372; FYMHYDRXLA.

DR TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.

DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.

DR Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron; Complete proteome.

KW METAL 144 144 IRON (POTENTIAL).

FT METAL 149 149 IRON (POTENTIAL).

FT SEQUENCE 289 AA; 33445 MW; 2D68931CE31DS21 CRC64;

Query Match 13.6%; Score 256.5; DB 1; Length 289;

Best Local Similarity 25.6%; Pred. No. 3.3e-13;

Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;

QY 106 LMVRLLSRSLWSKCYPRFLDYLAFAGLSLDFLDHQAVI-KFFELETHFSYVGV 164

DB 49 VVHELIITQQEVVTRACQAYDLGLANMLPTDRLPQLPEINRVLQRETQWQVEPVFALI 108

QY 165 APOHYLSLQDRYFFIASVMRTLDKCNFSLTLPDLTLDLGLHPVLLPSPSEFFINMGRLL 224

DB 109 SFDREFFALLADKKFVATFLRRREEFDYQEPDFFEHYVYGCAMLTHPDPFAAFTHYVQOL 168

QY 225 FTKVTKERKQVAPSKKRIQTLQSNLTAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQE 284

Db 169 GAKATPKERSY-----LARYWFTVFELGVORQQTKIYGGGILSSPGE 212

QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSTPQSTLFSIRHPDELVELTSK-----LEWM-LD 338

Db 213 TLVASESTIPKREPDFIMQVLRTPYRIDIMQPIYVLPDLSQLSORDVMALVQWQAMQ 272

QY 339 QGLLESIPLYNQEK 352

Db 273 DGLLP--PLFPQKE 284

RESULT 3

PH4H_PSEAE STANDARD; PRT; 262 AA.

AC P43334;

DT 01-NOV-1995 (Rel. 32, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-monooxygenase).

GN PH4H OR PA0872.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN NCBI_TaxID=287;

RP SEQUENCE FROM N.A.

RX STRAIN=ATCC 15692 / PA01;

RA MEDLINE=20437337; PubMed=8108417;

RA Zhao G., Xia L., Song J., Roy R.A.;

RT "Pseudomonas aeruginosa possesses homologues of mammalian phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DOH as part of a three-component gene cluster.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:1366-1370 (1994).

[2]

CC SEQUENCE FROM N.A.

CC STRAIN=ATCC 15692 / PA01;

CC MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Coulter S.N., Tolentino E., Westbrook-Wadman S., Yuan Y., Brody L.L., Folger K.R., Kas A., Larbig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";

RL Nature 406:959-964 (2000).

CC -!- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) = L-tyrosine + dihydrobiopterin + H(2)O.

CC -!- COFACTOR: Binds 1 ferrous ion.

CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.

CC -!- SUBUNIT: MONOMER.

CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M88627; AAA25936.1; --

DR EMBL; AE004522; AAG04261.1; --

DR HSSP; P04177; 1TOH.

DR InterPro; IPR001273; Aaa_hydroxylase.

DR Pfam; PF00351; bioprotein_H_1.

DR PRINTS; PR00372; FYMHYDRXLA.

DR TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.

DR PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.

DR Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron; Complete proteome.

KW


```

QY 159 PVSGFVAPHOYLSLQDRFPFIASVWRTLDKDNFSLTDLHLGHVPMILHPSSEFF 218
D 283 PVKGLSLARDFLASLAFRFVQCTOYVRHASSPMHSPEDCCCHLGHVPMILADKTPAQS 342
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSNLIAIVRCFMFTVESGLIENHGRKAYGAVL 278
D 343 QDIG-----LASIGATDEIEKTLATL-----WFTVVEGLCKRONGIYVAGAGL 386
QY 279 ISSPQELGAHFDINVRVLPLELDQILRLPNTSTPQETLFSIRHFEDE 325
D 387 LSSYGLHSLHSIDSEPEVRDPDPMAAVPCQDPQVYPVFSSEFSD 433

RESULT 7
TRSH RAT STANDARD; PRT; 444 AA.
AC P09810;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Typtophan 5-monooxygenase (EC 1.14.16.4) (Typtophan 5-hydroxylase).
GN Trp.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Pineal gland;
RX MEDLINE=88244702; PubMed=3379411;
RA Darnon M.C., Guilbert B., Leviet V., Ehret M., Maitre M., Mallet J.;
RT "Sequence of two mRNAs encoding active rat typtophan hydroxylase.";
RL J. Neurochem. 51:312-316(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91245924; PubMed=1645430;
RA Kim K.S., Wesel T.C., Stone D.M., Carver C.H., Joh T.H., Park D.H.;
RT "Molecular cloning and characterization of cDNA encoding typtophan
RT hydroxylase from rat central serotonergic neurons.";
RL Brain Res. Mol. Brain Res. 9:277-283(1991).
RN [3]
RP SEQUENCE OF 167-261 FROM N.A.
RX MEDLINE=87005247; PubMed=2875901;
RA Darnon M.C., Grima B., Cash C.D., Maitre M., Mallet J.;
RT "Isolation of a rat pineal gland cDNA clone homologous to tyrosine
RT and phenylalanine hydroxylases.";
RL FEBS Lett. 206:43-46(1986).
RL
CC -1- CATALYTIC ACTIVITY: L-typtophan + tetrahydropteridine + O(2) = 5-
CC hydroxy-L-typtophan + dihydropteridine + H(2)O.
CC
CC -1- COFACTOR: FERROUS ION.
CC
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC
CC -1- SIMILARITY: BELONGS TO THE BIPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC EMBL: M28000; AAA2262.1; -
CC EMBL: X53501; AAA37579.1; -
CC PIR: J10034; WHTW
CC PIR: A24367; A24367.
CC PIR: A60034; A60034.
CC HSP: P04177; I10H
CC Interpro: IPR002912; ACT.

```

```

DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; Bioplerin_H; 1.
DR Pfam: PF01842; ACT; 1.
DR PRINTS: PRO0372; PVMHYDRLASE.
DR TIGRfam: TIGR01270; TYP_5_monoox; 1.
DR PROSITE: PS00367; BIPTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY PRA) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
SQ SEQUENCE 444 AA; 51068 MW; C3CF5245727C825 CRC64;

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 1,le-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYLLSSRSFIMKSCPRFFIDYLEAFGLSDPDH-----QAVIFPELEHFGY 158
D 174 WGITRELMKLPTFACR---EILRNPLSKYIGYREDNVPLEVSNFKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLQDRFPFIASVWRTLDKDNFSLTDLHLGHVPMILHPSSEFF 218
D 231 PVAGYLSPPDFSLGAFRFVQCTOYVRHSSDPLTTPEDCHLGHVPLAEPSPAQS 290
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSNLIAIVRCFMFTVESGLIENHGRKAYGAVL 278
D 291 QDIG-----LASIGATSEVQKLTAT-----CYFTVVEGLCKQDQGLRVFGAGL 334
QY 279 ISSPQELGAHFDINVRVLPLELDQILRLPNTSTPQETLFSIRHFEDE 328.
D 335 LSSISLRLHSLGSHAKVFPDPKVAQCECLITSFDVYFVSEFSDAKE 384

RESULT 8
TRSH MOUSE STANDARD; PRT; 447 AA.
AC P15532;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Typtophan 5-monooxygenase (EC 1.14.16.4) (Typtophan 5-hydroxylase).
GN Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90243261; PubMed=2110547;
RA Scoll J., Kozak C.A., Goldman D.;
RT "Characterization and chromosomal mapping of a cDNA encoding
RT typtophan hydroxylase from a mouse mastocytoma cell line.";
RL Genomics 7:88-96(1990).
RL
CC -1- CATALYTIC ACTIVITY: L-typtophan + tetrahydropteridine + O(2) = 5-
CC hydroxy-L-typtophan + dihydropteridine + H(2)O.
CC
CC -1- COFACTOR: FERROUS ION.
CC
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC
CC -1- SIMILARITY: BELONGS TO THE BIPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC

```

```

DR EMBL; J04758; AAA63401.1; -.
DR PIR; A34582; A34582.
DR HSSP; P04177; ITOH.
DR MGI; MGI:98796; Tph.
DR InterPro; IPR002912; ACT.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FWHYDRXLASE.
DR TIGRFS; TIGR01270; Ttp 5 mono.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
DR Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 61 61 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT METAL 275 275 IRON (BY SIMILARITY).
FT METAL 280 280 IRON (BY SIMILARITY).
FT METAL 320 320 IRON (BY SIMILARITY).
SQ SEQUENCE 447 AA; 51343 MW; 16C839F22A138BCA CRC64;

Query Match 12.0%; Score 226.5; DB 1; Length 447;
Best Local Similarity 27.4%; Pred. No. 1.3e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSIMKSYCPFRFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
DB 177 WGTIFRELKLYPTHACR---EYLKPLLSKYCGYREDNIPQLEDVSNFLKERTGFSIR 233
QY 159 PVSGFVAPHOVLSDQRYFPPIASVNRITLDKNFSLTPDLHDLLGHVPMLLHPSSEFF 218
DB 234 PVAGYSPDFLSGLAFRVFCHCTQYVVRHSSDPYTPEDTCHELLGHVPLLAESPFAQFS 293
QY 219 INMGRLFTKVIKQVALPSKKORIOTLQSNLTAIVRCFWFTVSGLIENHEGRKAYGAVL 278
DB 294 QEIG-----LASLGASEEAVOKLAT-----CYFFTVFGLCKQDQQLRVFGAGL 337
QY 279 ISSQELGHAFIDNVRVLPDLQIIRLPFTNSTPQETLPSIRHFDLVE 328
DB 338 LSSISELKHLSGHAKVKPPDKPKIACKQELITSPQDYVFSSEFDEAKE 387

RESULT 9
TRSH RABIT STANDARD; PRT; 444 AA.
AC P17290; 029523;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tryptophan 5-monoxygenase (EC 1.14.16.4) (Tryptophan 5-hydroxylase).
GN TPH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87289639; PubMed=3475690;
RA Grenett H.E., Ledley F.D., Reed L.L., Woo S.L.C.;
RT Full-length cDNA for rabbit tryptophan hydroxylase: functional
RL domains and evolution of aromatic amino acid hydroxylases.;
RN Proc. Natl. Acad. Sci. U.S.A. 84:5530-5534 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95077422; PubMed=7986090;
RA Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;
RT "Cloning and expression of rabbit and human brain tryptophan
RL hydroxylase cDNA in Escherichia coli.";
RN Arch. Biochem. Biophys. 315:445-453 (1994).
CC -!- CATALYTIC ACTIVITY: L-tryptophan + tetrahydropteridine + O(2) = 5-
CC hydroxy-L-tryptophan + dihydropteridine + H(2)O.
CC -!- COPACITOR; FERROUS ION.
CC -!- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
CC SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST

```

```

CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -!- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M17250; AAA31487.1; -.
CC EMBL; L29305; AAA67051.1; -.
CC PIR; A32699; A32699.
CC HSSP; P04177; ITOH.
CC InterPro; IPR002912; ACT.
CC InterPro; IPR001273; Aaa hydroxylase.
CC Pfam; PF00351; bioterin_H; 1.
CC Pfam; PF01842; ACT; 1.
CC PRINTS; PR00372; FWHYDRXLASE.
CC TIGRFS; TIGR01270; Ttp 5 mono.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
FT METAL 317 317 IRON (BY SIMILARITY).
FT CONFLICT 102 102 M -> L (IN REF. 1).
FT CONFLICT 151 151 L -> S (IN REF. 2).
FT CONFLICT 202 203 KY -> ND (IN REF. 1).
FT CONFLICT 207 207 R -> Q (IN REF. 2).
FT CONFLICT 390 390 T -> K (IN REF. 1).
SQ SEQUENCE 444 AA; 51118 MW; BF182451B28ECD80 CRC64;

Query Match 11.9%; Score 225.5; DB 1; Length 444;
Best Local Similarity 27.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 44; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSIMKSYCPFRFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
DB 174 WGTIFRELKLYPTHACR---EYLKPLLSKYCGYREDNIPQLEDVSNFLKERTGFSIR 230
QY 159 PVSGFVAPHOVLSDQRYFPPIASVNRITLDKNFSLTPDLHDLLGHVPMLLHPSSEFF 218
DB 231 PVAGYSPDFLSGLAFRVFCHCTQYVVRHSSDPYTPEDTCHELLGHVPLLAESPFAQFS 290
QY 219 INMGRLFTKVIKQVALPSKKORIOTLQSNLTAIVRCFWFTVSGLIENHEGRKAYGAVL 278
DB 291 QEIG-----LASLGASEEAVOKLAT-----CYFFTVFGLCKQDQQLRVFGAGL 334
QY 279 ISSQELGHAFIDNVRVLPDLQIIRLPFTNSTPQETLPSIRHFDLVE 328
DB 335 LSSISELKHLSGHAKVKPPDKPKIYKQELITFPQDYVFSSEFDEAKE 384

RESULT 10
TY3H ANGAN STANDARD; PRT; 488 AA.
ID TY3H ANGAN
AC 042091;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monoxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
GN TH.
OS Anguilla anguilla (European freshwater eel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;
OC Anguilla
OX NCBI_TaxID=7936;
RN [1]

```



```

RE SEQUENCE FROM N.A.
RX MEDLINE=98344760; PubMed=9681435;
RA Bouliand S., Biguet N.F., Vidal B., Veron M., Mallet J.,
RT Vincent J.D., Ducour S., Vernier P.,
RT "Tyrosine hydroxylase in the european eel (Anguilla anguilla): cDNA
RT cloning, brain distribution, and phylogenetic analysis.";
RL J. Neurochem. 71:460-470(1998).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: Catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ000731; CA04264.1; -.
CC HSP; P04177; ITOH.
CC InterPro: IPR001273; Aaa_hydroxylase.
CC Pfam; PF00351; bioperin_H; 1.
CC PRINTS; PR00372; FYMHDXLXASE.
CC TIGRFAMs; TIGR01269; Tyr_3_monoox; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC KEGG; catcholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 38 38 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
CC FT METAL 321 321 IRON (BY SIMILARITY).
CC FT METAL 326 326 IRON (BY SIMILARITY).
CC FT METAL 366 366 IRON (BY SIMILARITY).
CC SO SEQUENCE 488 AA; 55490 MW; 57353BA39EBD448 CRC64;

Query Match 11.9%; Score 225.5; DB 1; Length 488;
Best Local Similarity 28.1%; Pred. No. 1.8e-10;
Matches 65; Conservative 38; Mismatches 99; Indels 29; Gaps 5;

OY 107 MYRLSSRFSLMKSVCPRFFLDYLAFLGSLDFLDH-----QAVIKFELETHFSY 157
DB 223 WREYVSTLRDLTYTHACS-----EHLERFLRERHCGYSPNSIPQLEVDVSHLKERITGFOL 278
OY 158 YPVSGFVAAPHQVYSLDODRFYPLASVMTLKDGNFSLPFDLIHLGHPVPLHPSFSEF 217
DB 279 RPYAGLLSARDPLASLAFRFQCTGYIRASSPMSPEPDCHEHLGHPVPLADRTAQAQF 338
OY 218 FINNGRLFTVIEQVALPSKQRIOTLOSNIALVRCFMTFVSGLIENHGRKAYGAV 277
DB 339 SQNG-----LASLGASEEDIKST-----LYWTFVEGLCKQDGVKAYGAG 382
OY 278 LISSPDELGHAFIDNVRVPLELDQIIRLPNTSTPQETLSIRHFDLVE 328
DB 383 LLSSYGLVHSLSDPERRERFDEAAAPYODVQSYVFSSESFTDAKE 433

RESULT 11
TY3H MOUSE STANDARD; PRT; 498 AA.
AC P24529;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RX SEQUENCE FROM N.A.
RA MEDLINE=91248263; PubMed=1674869;
RA Ichikawa S., Sasaka T., Nagatsu T.;
RT "Primary structure of mouse tyrosine hydroxylase deduced from its
RT cDNA.";
RL Biochem. Biophys. Res. Commun. 176:1610-1616(1991).
CC [2]
CC SEQUENCE OF 1-30 FROM N.A.
CC STRAIN=BALE/cj
RA Morgan W.W., Bernudez J., Sharp Z.D.;
RL Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: Catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69200; AAA0434.1; -.
CC PIR; JN0068; JN0068.
CC HSP; P04177; ITOH.
CC MGD; MGI:98735; Th.
CC InterPro: IPR001273; Aaa_hydroxylase.
CC Pfam; PF00351; bioperin_H; 1.
CC PRINTS; PR00372; FYMHDXLXASE.
CC TIGRFAMs; TIGR01269; Tyr_3_monoox; 1.
CC PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
CC KEGG; catcholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 19 19 PHOSPHORYLATION (BY CMK2) (BY
CC SIMILARITY)
CC FT MOD_RES 31 31 PHOSPHORYLATION (BY SIMILARITY).
CC FT DOMAIN 51 59 POLY-ALA
CC FT METAL 331 331 IRON (BY SIMILARITY).
CC FT METAL 336 336 IRON (BY SIMILARITY).
CC FT METAL 376 376 IRON (BY SIMILARITY).
CC SO SEQUENCE 498 AA; 55992 MW; 62790179664F6DC6 CRC64;

Query Match 11.9%; Score 225.5; DB 1; Length 498;
Best Local Similarity 27.5%; Pred. No. 1.8e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;

OY 107 MYRLSSRFSLMKSVCPRFFLDYLAFLGSLDFLDH-----QAVIKFELETHFSY 158
DB 223 WREYVSTLRDLTYTHACR-----EHLERFLRERHCGYREDISIPQLEVDVSHLKERITGFOL 289
OY 159 YPVSGFVAAPHQVYSLDODRFYPLASVMTLKDGNFSLPFDLIHLGHPVPLHPSFSEF 218
DB 290 PYAGLLSARDPLASLAFRFQCTGYIRASSPMSPEPDCHEHLGHPVPLADRTAQAQF 349
OY 219 FINNGRLFTVIEQVALPSKQRIOTLOSNIALVRCFMTFVSGLIENHGRKAYGAVL 278
DB 350 QDIE-----LASLGASEDEIKST-----YVWTFVEGLCKQDGVKAYGAGL 393
OY 279 LISSPDELGHAFIDNVRVPLELDQIIRLPNTSTPQETLSIRHFDLVE 324

```


Db 394 LSSYGELLSLSBEPEVRAFPDPTAAVQPDQTYQPVYFVSEFSFSDAKLNNVASRIQ 453

RESULT 12

TR5H_XENLA STANDARD; PRT; 481 AA.
ID Q92142;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Tryptophan 5-monooxygenase (EC 1.14.16.4) (tryptophan 5-hydroxylase).
GN Tph.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=94246419; PubMed=8189245;
RA Green C.B., Besharse J.C.;
RT "Tryptophan hydroxylase expression is regulated by a circadian clock
in Xenopus laevis retina.";
RL J. Neurochem. 62:2420-2428(1994).
CC -|- CATALYTIC ACTIVITY: L-tryptophan + tetrahydropteridine + O(2) = 5-
hydroxy-L-tryptophan + dihydropteridine + H(2)O.
CC -|- COPACTOR: FERROUS ION.
CC -|- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -|- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L20679; AAA21306.1; --
HSSP; P04177; ITOH.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001273; Aaa hydroxylase.
DR Pfam; PF00351; bioterin_H; 1.
DR Pfam; PF01842; ACT; 1.
DR PRINTS; PR00372; FWHYDXYLASE.
DR TIGRfam; TIGR01270; TTD_5_MONOXYL; 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron.
FT METAL 309 309 IRON (BY SIMILARITY).
FT METAL 314 314 IRON (BY SIMILARITY).
FT METAL 354 354 IRON (BY SIMILARITY).
SQ SEQUENCE 481 AA; 55406 MW; 6852C33EFF8DEBA0 CRC64;

Query Match 11.8%; Score 223.5; DB 1; Length 481;
Best Local Similarity 29.5%; Pred. No. 2.5e-10;
Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;

QY 107 WYRLSSRFLSKSYCPFRFLDYLAFLGLSLDFLDH-----QAVIKPELETHFSY 158
Db 211 WGTFRRLNKLPHYTHACK-----EYLNKLPLSKHGYRDNIPOLEDVSRFLRBTGFTIR 267
QY 159 PVSGFVAFPHYLSLQDRYFPIASVNRITLKDKNFSL-TPDLIHLLGHVPMLLHPFSFSE 217
Db 268 PVAGVLSPRDLAFLAFVHCYVVR-DSPLNTPEDTCHELLGHVPLLAEPSPFAQ 326
QY 218 FNNGLFTKVIKVOALPSKQRIQTLQSLNLAIVRCFWETVESGLIENHGRKAYGAV 277
Db 327 SQETG-----LASLGASDEAVQKLAT-----CYFFTFVGLCKQEGKLKVGAG 370

QY 278 LISSPOELGHAFIDNVRLPVLDELQIIRLPNTSTPQRTLFSIRHFDLVELTSLKLE 334
Db 371 LLSSISELKHSLSGNAKVPFD-----PMTCN-QBCI--ITSFOELYFVSESE 417

RESULT 13

TY3H_RAT STANDARD; PRT; 498 AA.
ID P04177;
AC 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (tyrosine 3-hydroxylase) (TH).
GN TH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85113249; PubMed=2857492;
RA Grima B., Lamouroux A., Blanot F., Faucon Biquet N., Mallet J.;
RT "Complete coding sequence of rat tyrosine hydroxylase mRNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:617-621(1985).
CC [2]
RP SEQUENCE FROM N.A.
RA Anton X.X., Manaster J.S., Kordower X.X., Markham X.X., Bredeesen D.E.;
RN Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
CC [3]
RP PHOSPHORYLATION
RX MEDLINE=91170235; PubMed=1672315;
RA Haycock J.W., Haycock D.A.;
RT "Tyrosine hydroxylase in rat brain dopaminergic nerve terminals.
Multiple-site phosphorylation in vivo and in synaptosomes.";
RL J. Biol. Chem. 266:5650-5657(1991).
CC [4]

X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 164-498.
MEDLINE=97372896; PubMed=9228951;
RA Goodwill K.E., Sabatier C., Marks C., Raag R., Fitzpatrick P.F.,
RA Stevens R.C.;
RT "Crystal structure of tyrosine hydroxylase at 2.3 A and its
implications for inherited neurodegenerative diseases.";
RT Nat. Struct. Biol. 4:578-585(1997).
CC -|- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
NEURONES.
CC -|- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -|- COPACTOR: FERROUS ION.
CC -|- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CATALYTIC ACTIVITY.
CC -|- PATHWAY: Catecholamine biosynthesis; first step.
CC -|- SUBUNIT: HOMOTETRAMER.
CC -|- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M10244; AAA42257.1; --
EMBL; L22651; AAA42258.1; --
DR PIR; A00510; WHRTY.
DR PDB; 1TOH; 03-JUN-98.
DR InterPro; IPR001273; Aaa hydroxylase.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FWHYDXYLASE.
DR TIGRfam; TIGR01269; TYX_3_MONOXYL; 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.

KW Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
 RN Neurotransmitter biosynthesis; Phosphorylation (by CAMK2).
 FT MOD_RES 19 19 PHOSPHORYLATION (BY CAMK2).
 FT MOD_RES 31 31 PHOSPHORYLATION (BY PKA).
 FT MOD_RES 40 40 PHOSPHORYLATION (BY PKA).
 FT DOKAIN 51 59 POLY-ALA.
 FT METAL 331 331 IRON.
 FT METAL 336 336 IRON.
 FT METAL 376 376 IRON.
 SQ SEQUENCE 498 AA; 55965 MW; 17F7E003D29218C5 CRC64;

Query Match 11.8%; Score 223; DB 1; Length 498;
 Best Local Similarity 26.7%; Pred. No. 2.9e-10;
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRSFLMS-----YCRPFLDYLEAFGLSDPLDH-----QAVIKPELETHF 155
 DB 227 ABEIATNKEVYVTLKGLYATHACREHESFQLEKRCYGRSDISTPLEDVSRLKERTGF 286
 QY 156 SYPSGFAVPHQYLSLQDRYPPIASVWRTLDKDNFSLTPDLIDLGHVPMILHPSFS 215
 DB 287 QLRVAGLLSRDPLASLAFRVQCTQYIRHASSPMHSPBPCCHLGHVPLADRTFA 346
 QY 216 EFTINMGLFTKVIKYOALPSKKORIQTGLSNLIAVRCFMFTVESGLIENHGRKAYG 275
 DB 347 QFSODIG-----LASIGASDEIEKLTST-----VWFTEFGLCKONGELKAYG 330
 QY 276 AVLSPOELGHAFIDNVRVLPLELDQIIRLPFNTSTPQETLFSIRHF-----DELVELTS 331
 DB 391 AGLSSYGEHLHSISEPEVRAPDPTAAVQPYODQYQVYFVSESFNDAKDLRYNYS 450
 QY 332 KLE 334
 DB 451 RIQ 453

RESULT 14
 ID TY3H BOVIN STANDARD; PRT; 490 AA.
 AC p12789;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
 GN Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88259287; PubMed=2898537;
 RA D'Wello S.R., Weisberg E.P., Stachowiak M.K., Turzai L.M., Gloio A.E.,
 RA Kaplan B.B.;
 RT "Isolation and nucleotide sequence of a cDNA clone encoding bovine
 RT adrenal tyrosine hydroxylase: comparative analysis of tyrosine
 RT hydroxylase gene products.";
 RN J. Neurosci. Res. 19:440-449 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274405; PubMed=2899135;
 RA Saadat S., Stehle A.D., Lamouroux A., Mallet J., Thoenen H.;
 RT "Predicted amino acid sequence of bovine tyrosine hydroxylase and its
 RT similarity to tyrosine hydroxylases from other species.";
 RN J. Neurochem. 51:572-578 (1988).
 RN [3]
 RP SEQUENCE OF 153-169.
 RX MEDLINE=88183482; PubMed=2895648;
 RA Abate C., Smith J.A., Jon T.H.;
 RT "Characterization of the catalytic domain of bovine adrenal tyrosine
 RT hydroxylase.";
 RN Biochem. Biophys. Res. Commun. 151:1446-1453 (1988).

RN [4]
 RP SEQUENCE OF 1-27.
 RC TISSUE=adrenal medulla;
 RX MEDLINE=88163796; PubMed=2894860;
 RA Haavik J., Anderson K.K., Petersen L., Flatmark T.;
 RT "Soluble tyrosine hydroxylase (tyrosine 3-monooxygenase) from bovine
 RT adrenal medulla: large-scale purification and physicochemical
 RT properties.";
 RN Biochim. Biophys. Acta 953:142-156 (1988).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
 CC NEURONES.
 CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
 CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
 CC -1- COFACTOR: FERROUS ION.
 CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
 CC CATALYTIC ACTIVITY.
 CC -1- PATHWAY: catecholamine biosynthesis; first step.
 CC -1- SIMILARITY: BELONGS TO THE BIOTRIN-DEPENDENT AROMATIC AMINO ACID
 CC HYDROXYLASES FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M36794; AAA30779.1; -;
 DR EMBL; M36705; AAA30798.1; -;
 DR PIR; A27687; A27687.
 DR PIR; J10039; J10039.
 DR HSSP; P04177; IYOH.
 DR InterPro: IPR001273; Aaa hydroxylase.
 DR Pfam: PF00351; bioperin H; 1.
 DR PRINTS: PR00372; FYMHYDXLASE.
 DR TIGRfam: TIGR01269; Tyr 3 monoox.
 DR PROSITE: PS00367; BIOTRIN HYDROXYL.
 DR Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
 KW Neurotransmitter biosynthesis; Phosphorylation.
 RN [1]
 RP INIT MET 0
 FT MOD_RES 39 39 PHOSPHORYLATION (BY PKA).
 FT METAL 323 323 IRON (BY SIMILARITY).
 FT METAL 328 328 IRON (BY SIMILARITY).
 FT METAL 368 368 IRON (BY SIMILARITY).
 FT CONFLICT 64 67 AAWL -> GSLV (IN REF. 2).
 FT CONFLICT 72 72 E -> K (IN REF. 2).
 FT CONFLICT 82 82 R -> R (IN REF. 2).
 FT CONFLICT 85 85 R -> V (IN REF. 2).
 FT CONFLICT 283 283 A -> V (IN REF. 2).
 FT CONFLICT 320 320 E -> D (IN REF. 2).
 FT CONFLICT 327 329 GHV -> AHG (IN REF. 2).
 FT CONFLICT 379 379 K -> N (IN REF. 2).
 FT CONFLICT 470 470 H -> R (IN REF. 2).
 SQ SEQUENCE 490 AA; 54992 MW; 5F54PF23C0EABD CRC64;

Query Match 11.7%; Score 221.5; DB 1; Length 490;
 Best Local Similarity 27.1%; Pred. No. 3.7e-10;
 Matches 65; Conservative 42; Mismatches 102; Indels 31; Gaps 5;

QY 107 WYRLSSRSFLMSKYCPFLDYLEAFGLSDPLDH-----QAVIKPELETHFSY 158
 DB 225 WKVEYTLKGLYATHACREHESFQLEKRCYGRSDISTPLEDVSRLKERTGF 281
 QY 159 PVSGFAVPHQYLSLQDRYPPIASVWRTLDKDNFSLTPDLIDLGHVPMILHPSFSFE 218
 DB 282 PAAGLLSARDPLASLAFRVQCTQYIRHASSPMHSPBPCCHLGHVPLADRTFA 341
 QY 219 INNGRLFTKVIKYOALPSKKORIQTGLSNLIAVRCFMFTVESGLIENHGRKAYG 278
 DB 342 QDIG-----LASIGASDEIEKLTST-----LYWTFEFGCKONGELKAYG 385
 QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNTSTPQETLFSIRHF-----DELVELTSKLE 334


```

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: catecholamine biosynthesis; first step.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00414; CAA68472.1; ALT_SEQ.
DR EMBL: M17588; AAA61179.1; -.
DR EMBL: X05290; CAA28908.1; -.
DR EMBL: M24791; AAA61173.1; -.
DR EMBL: M24787; AAA61173.1; JOINED.
DR EMBL: M24788; AAA61173.1; JOINED.
DR EMBL: M24791; AAA61170.1; JOINED.
DR EMBL: M24787; AAA61170.1; JOINED.
DR EMBL: M20911; AAA61167.1; -.
DR PIR: A27791; WHH01.
DR PIR: A27791; WHH02.
DR PIR: A27791; WHH03.
DR PIR: A26825; WHH04.
DR PIR: JE0012; JE0012.
DR PIR: JE0013; JE0013.
DR PIR: JE0014; JE0014.
DR HSSP: P04177; IYOH.
DR Genew: HGNC:11782; TH.
DR MIM: 191290; -.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; bioplerin_H; 1.
DR PRINTS: PR00372; FYMHYDRXLASE.
DR TIGRFAMs: TIGR01269; Tyr_3_monoox; 1.
DR PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW alternative splicing; Neurotransmitter biosynthesis; Phosphorylation.
KW MOD RES 71 71 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT DOMAIN 85 90 POLY-ALA.
FT METAL 361 361 IRON (BY SIMILARITY).
FT METAL 366 366 IRON (BY SIMILARITY).
FT METAL 406 406 IRON (BY SIMILARITY).
FT VARSPIC 1 33 MISSING (IN ISOFORM 4).
FT VARSPIC 34 34 O -> M (IN ISOFORM 4).
FT VARSPIC 31 61 MISSING (IN ISOFORM 1).
FT VARSPIC 35 61 MISSING (IN ISOFORM 2).
SQ SEQUENCE 528 AA; 58524 MW; B3DC349955A21074 CRC64;

Query Match 11.6%; Score 220; DB 1; Length 528;
Best Local Similarity 26.7%; Pred. No. 5, 3e-10;
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

```

```

QY 276 AVIISPDEIIGHAFIDNVRVLELEIDQITRPNTSTPQETLSIRNF-----DELVELTS 331
DB 421 AGLLSYGELHCLSEPEIRAFDEAAVOPDQTYGVSVFSSPSDADKLSVNS 480
QY 332 KLE 334
DB 481 RIQ 483

RESULT 17
ID TY3H DROME STANDARD; PRT; 508 AA.
AC P18459;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH
DE (Protein Pale).
GN TH OR PLE.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A. PubMed=2483109;
RX MEDLINE=90165583;
RA Neckmeyer W.S., Quinn M.G.,
RT "Isolation and characterization of the gene for Drosophila tyrosine
RT hydroxylase".
RU Neuron 2:1167-1175(1989).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: catecholamine biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14395; AAA62877.1; -.
DR EMBL: X76209; CAA53802.1; -.
DR PIR: JN0010; JN0010.
DR HSSP: P04177; IYOH.
DR FLYBase: FBgn0005626; ple.
DR InterPro: IPR001273; Aaa_hydroxylase.
DR Pfam: PF00351; bioplerin_H; 1.
DR PRINTS: PR00372; FYMHYDRXLASE.
DR TIGRFAMs: TIGR01269; Tyr_3_monoox; 1.
DR PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
KW MOD RES 338 338 IRON (BY SIMILARITY).
FT METAL 343 343 IRON (BY SIMILARITY).
FT METAL 383 383 IRON (BY SIMILARITY).
SQ SEQUENCE 508 AA; 57861 MW; 2D194E08455E6F CRC64;

Query Match 11.6%; Score 219; DB 1; Length 508;
Best Local Similarity 28.6%; Pred. No. 6, 1e-10;
Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;

```

```

QY 180 IASVMTLDKDNFSLPDTHLLGHVPLHLPSPSEFFINMGLFTKVKIEKQVALPSKK 239
DB 318 STQVVRHNSPYHTPEPDSHELLGHMPLADSPSPQSQEIG-----LASLGASD 368
QY 240 QRIOTLOSNIIVRCFWFTVFSGLIENHEGRKAYGAVLISSPOELGHAFIDNRVPLE 299
DB 369 EEIEKIST-----VWTFVFLGCKEHQKATKAGGLSSYCELLHAISDKCEHRAFE 421
QY 292 --NVRVLPLE-----LDQIIRLPNTSTPOETLFSIRHFD--ELVELT 330
DB 422 PASTAVOYQDQEQYQPIYYVAESFEDAKDKFRWVTSWTSRFEVRFN-PHTEVEVLDV 480
QY 331 SKLEWMLDQGLLESIPLYN 349
DB 481 DKLETVHQMTTEILHUTN 499

RESULT 18
TY3H SCHMA
ID TY3H SCHMA STANDARD; PRT; 465 AA.
AC O17446;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tyrosine 3-monoxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
GN TH.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Puerto Rican;
RX MEDLINE=98421787; PubMed=9751167;
RA Hamdan F.F., Ribeiro P.;
RT "Cloning and characterization of a novel form of tyrosine hydroxylase
RT from the human parasite, Schistosoma mansoni.";
RL J. Neurochem. 71:1369-1380(1998).
CC -!- CATALYTIC ACTIVITY: L-tyrosine + tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-L-phenylalanine + dihydropteridine + H(2)O.
CC -!- COFACTOR: FERROUS ION.
CC -!- PATHWAY: Catecholamine biosynthesis; first step.
CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DB EMBL; AF030336; AAC62256.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; biotpterin_H; 1.
DR TIGRFAMs; TIGR01269; Tyr_3_monoox; 1.
DR PROSITE; PS00367; BIOTPTERIN_HYDROXYL; 1.
KW Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
FT METAL 294 294 IRON (BY SIMILARITY).
FT METAL 299 299 IRON (BY SIMILARITY).
FT METAL 339 339 IRON (BY SIMILARITY).
SQ SEQUENCE 465 AA; 54081 MW; 0DE47B4A474665CB CRC64;

Query Match 11.5%; Score 216.5; DB 1; Length 465;
Best Local Similarity 24.9%; Pred. No. 8.6e-10;
Matches 59; Conservative 49; Mismatches 88; Indels 41; Gaps 5;

QY 128 DYLEAFGLSDFLDH-----QAVIKFELETHFSYYPVSGFVAPHVLSLLQDRYFP 179

```

```

DB 214 EYIDGOLFLEKYNSESIPOLOTICEPMHRTSGFRIRPVAGLVSPKDFLASLAFRVQ 273
QY 180 IASVMTLDKDNFSLPDTHLLGHVPLHLPSPSEFFINMGLFTKVKIEKQVALPSKK 239
DB 274 CTQIRHRSRPMTPPDCHELIGHMPLVNRQFADFSQELG-----LASLGASEET 327
QY 240 QRIOTLOSNIIVRCFWFTVFSGLIENHEGRKAYGAVLISSPOELGHAFIDNRVPLE 299
DB 328 TRLSTL-----YWTFVFLGCKENETGALGAGIMSSYGELENAF----- 367
QY 300 LDQIIRLPNTSTPOETLFSIRHFDVELTSKLEWM-----LDQGLLESIPLYN 349
DB 368 SLDGVKPEFFNIDAAVQVYDDVGQYQKIYFVTESIESMKRELNRNYINTSGKSTPIYD 424

RESULT 19
PH4H RAT
ID PH4H RAT STANDARD; PRT; 453 AA.
AC P04176;
DT 20-MAR-1987 (Rel. 04, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
DE monooxygenase).
GN PAH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86140234; PubMed=2869038;
RA Dahl H.-H.M., Mercer J.F.B.;
RT "Isolation and sequence of a cDNA clone which contains the complete
RT coding region of rat phenylalanine hydroxylase. Structural homology
RT with tyrosine hydroxylase, glucocorticoid regulation, and use of
RT alternate polyadenylation sites.";
RL J. Biol. Chem. 261:4148-4153(1986).
RN [2]
RP SEQUENCE OF 208-453 FROM N.A.
RX MEDLINE=85122617; PubMed=6098294;
RA Robson K.J.H., Beattie W., James R.J., Cotton R.C.H., Morgan P.J.,
RA Woo S.L.C.;
RT "Sequence comparison of rat liver phenylalanine hydroxylase and its
RT cDNA clones";
RL Biochemistry 23:5671-5675(1984).
RN [3]
RP SEQUENCE OF 12-21 FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=80220293; PubMed=7387651;
RA Wretborn M., Humble E., Ragnarsson U., Engstrom L.;
RA Wretborn M., Humble E., Ragnarsson U., Engstrom L.;
RT "Amino acid sequence at the phosphorylated site of rat liver
RT phenylalanine hydroxylase and phosphorylation of a corresponding
RT synthetic peptide.";
RL Biochem. Biophys. Res. Commun. 93:403-408(1980).
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=99260738; PubMed=10331871;
RA Kobe B., Jennings I.G., House C.M., Michell B.J., Goodwill K.E.,
RA Santasiero B.D., Stevens R.C., Cotton R.G., Kemp B.E.;
RT "Structural basis of autoregulation of phenylalanine hydroxylase.";
RL Nat. Struct. Biol. 6:442-448(1999).
CC -!- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
CC L-tyrosine + dihydrobiopterin + H(2)O.
CC -!- COFACTOR: FERROUS ION.
CC -!- ENZYME REGULATION: N-TERMINAL REGION OF PAH IS THOUGHT TO CONTAIN
CC ALLOSTERIC BINDING SITES FOR PHENYLALANINE AND TO CONSTITUTE AN
CC "INHIBITORY" DOMAIN THAT REGULATES THE ACTIVITY OF A CATALYTIC
CC DOMAIN IN THE C-TERMINAL PORTION OF THE MOLECULE.
CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.

```

```

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: M12337; AAA41843.1; -
DR      EMBL: K02599; AAA41794.1; -
DR      PIR: A00509; WHTRF.
DR      PIR: A14970; A14970.
DR      PIR: A25321; A25321.
DR      PDB: 1PHZ; 30-APR-99.
DR      PDB: 2PHM; 30-APR-99.
DR      InterPro: IPR002912; ACT.
DR      InterPro: IPR001273; Aaa_hydroxylase.
DR      Pfam: PF01842; ACT.1.
DR      Pfam: PF01842; ACT.1.
DR      PRINTS: PR00372; FWHYDRYLASE.
DR      TIGRfam: TIGR01268; PheHydrox_tetr; 1.
DR      PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
KW      Oxidoreductase; Monooxygenase; Phosphorylation;
KW      Phenylalanine catabolism; Iron; 3D-structure.
FT      MOD_RS 16 16 PHOSPHORYLATION (BY PKA).
FT      METAL 285 285 IRON.
FT      METAL 290 290 IRON.
FT      METAL 330 330 IRON.
SQ      SEQUENCE 453 AA; 51821 MW; 365D9E8A7E498D52 CRC64;
Query Match 11.3%; Score 214; DB 1; Length 453;
Best Local Similarity 28.4%; Pred. No. 1.3e-09;
Matches 61; Conservative 37; Mismatches 89; Indels 28; Gaps 4;
OY      115 FSLWKSYCPFLDYLEAFGLSDPLDH-OAVIKFPELETHPSYYPGAFAPHOYSL 173
DB      210 PFLLEKCY-----GFRDNI PQLEEDVSQFLQCTGPRAPVAGLLSMDPLGGL 258
OY      174 QDRYFPIASVMTLDNDNLSLPDLHDHLDLGHVPLWLPSPSEPFIMNGRLFTVIEKYG 233
DB      259 AFRVHCTQYIRHSGKRWYTPEDDICHLELGHVPLPFSDBSPAQSGRIE-----LA 309
OY      234 ALPSKKQRIOTLOSNIAYRCWFVPSGLIENHGKRAYCAVLISFQELGHAFIDNV 293
DB      310 SLGAPDYIEKLAT-----IYFVTEFGELCKEGDSIRKAYAGLLSFGELQICUSDKP 362
OY      294 RVLPFLDQIIRLPENTSTPQETLFSIRHDELVE 328
DB      363 KLPLLEKTAQCEYSVTEPQPLVVAESFSDAKE 397
RESULT 20
PH4H_HUMAN STANDARD; PRT; 452 AA.
AC      P00439; Q16717;
DT      21-JUL-1986 (Rel. 01, Created)
DT      21-JUL-1986 (Rel. 01, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
DE      monooxygenase).
GN      PAH.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RS      SEQUENCE FROM N.A.
RC      TISSUE=Liver;
RA      KMOX S.C.M., Ledley F.D., Dilella A.G., Robson K.J.H., Woo S.L.C.;
RT      "Nucleotide sequence of a full-length complementary DNA clone and
RT      amino acid sequence of human phenylalanine hydroxylase.";

```

```

RL      Biochemistry 24:556-561(1985).
RL      [2]
RL      SEQUENCE FROM N.A.
RL      RA Scriber C.R., Nowacki P.M., Byck S., Prevost L.;
RL      RA Submitted (SEP-1997) to the EMBL/GenBank/DBD databases.
RL      [3]
RL      X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 117-424.
RL      RA MEDLINE=98069646; PubMed=9406548;
RL      RA Erlandsen H., Fusetti F., Martinez A., Hough E., Flatmark T.,
RL      RA Stevens R.C.;
RL      RA "Crystal structure of the catalytic domain of human phenylalanine
RL      RT hydroxylase reveals the structural basis for phenylketonuria.";
RL      RL Nat. Struct. Biol. 4:995-1000(1997).
RL      [4]
RL      X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 117-424.
RL      RA MEDLINE=99060040; PubMed=9843368;
RL      RA Erlandsen H., Flatmark T., Stevens R.C., Hough E.;
RL      RA "Crystallographic analysis of the human phenylalanine hydroxylase
RL      RT catalytic domain with bound catechol inhibitors at 2.0-A resolution.";
RL      RL Biochemistry 37:15638-15646(1998).
RL      [5]
RL      X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 117-452.
RL      RA MEDLINE=98307935; PubMed=9642259;
RL      RA Fusetti F., Erlandsen H., Flatmark T., Stevens R.C.;
RL      RA "Structure of tetrameric human phenylalanine hydroxylase and its
RL      RT implications for phenylketonuria.";
RL      RL J. Biol. Chem. 273:16962-16967(1998).
RL      [6]
RL      REVIEW ON PKU VARIANTS.
RL      RA MEDLINE=91348681; PubMed=1679029;
RL      RA Konecki D.S., Lichter-Konecki U.;
RL      RA "The phenylketonuria locus: current knowledge about alleles and
RL      RT mutations of the phenylalanine hydroxylase gene in various
RL      RT populations.";
RL      RL Hum. Genet. 87:377-388(1991).
RL      [7]
RL      REVIEW ON PKU VARIANTS.
RL      RA MEDLINE=91061429; PubMed=2246658;
RL      RA Cotton R.G.;
RL      RA "Heterogeneity of phenylketonuria at the clinical, protein and DNA
RL      RT levels.";
RL      RL J. Inher. Metab. Dis. 13:739-750(1990).
RL      [8]
RL      REVIEW ON PKU VARIANTS.
RL      RA MEDLINE=93244626; PubMed=1301187;
RL      RA Eisenmith R.C., Woo S.L.C.;
RL      RA "Molecular basis of phenylketonuria and related
RL      RT hyperphenylalaninemia: mutations and polymorphisms in the human
RL      RT phenylalanine hydroxylase gene.";
RL      RL Hum. Mutat. 1:13-22(1992).
RL      [9]
RL      DATABASE OF PKU VARIANTS.
RL      RA MEDLINE=96174613; PubMed=8594560;
RL      RA Hoang L., Byck S., Prevost L., Scriber C.R.;
RL      RA "PAH Mutation Analysis Consortium Database: a database for disease-
RL      RT producing and other allelic variation at the human PAH locus.";
RL      RL Nucleic Acids Res. 24:127-131(1996).
RL      [10]
RL      VARIANT PRO-311.
RL      RA MEDLINE=88294030; PubMed=2840952;
RL      RA Lichter-Konecki U., Konecki D.S., Dilella A.G., Brayton K., Marvit J.,
RL      RA Hahn T.M., Trefz F.K., Woo S.L.C.;
RL      RA "Phenylalanine hydroxylase deficiency caused by a single base
RL      RT substitution in an exon of the human phenylalanine hydroxylase
RL      RT gene.";
RL      RL Biochemistry 27:2881-2885(1988).
RL      [11]
RL      VARIANT LYS-280;
RL      RA MEDLINE=89190664; PubMed=2564729;
RL      RA Lyonnet S., Caillaud C., Rey F., Bertelson M., Frezal J., Rey J.,
RL      RA Wunnich A.;
RL      RA "Molecular genetics of phenylketonuria in Mediterranean countries: a
RL      RT mutation associated with partial phenylalanine hydroxylase

```

RA deficiency.";
 RT Am. J. Hum. Genet. 44:511-517 (1989).
 RN [12]
 RP VARIANT PRO-311.
 RX MEDLINE=90136055; PubMed=2615649;
 RA Hofman K.J., Antonarakis S.B., Missiou-Tsangarakis S., Boehm C.D.,
 RA Valle D.;
 RT "Phenylketonuria in the Greek population. Haplotype analysis of the
 RT phenylalanine hydroxylase gene and identification of a PKU
 RT mutation.";
 RL Mol. Biol. Med. 6:245-250 (1989).
 RN [13]
 RP VARIANT LEU-364 DEL.
 RX MEDLINE=90368081; PubMed=1975559;
 RA Svensson E., Andersson B., Hagenfeldt L.;
 RT "Two mutations within the coding sequence of the phenylalanine
 RT hydroxylase gene.";
 RL Hum. Genet. 85:300-304 (1990).
 RN [14]
 RP VARIANT GLN-261.
 RX MEDLINE=91150775; PubMed=1671810;
 RA Dianzani I., Forrest S.M., Camaschella C., Saglio G., Ponzone A.,
 RA Cotton R.G.;
 RT "Screening for mutations in the phenylalanine hydroxylase gene from
 RT Italian patients with phenylketonuria by using the chemical cleavage
 RT method: a new splice mutation.";
 RL Am. J. Hum. Genet. 48:631-635 (1991).
 RN [15]
 RP VARIANT SER-255.
 RX MEDLINE=91196738; PubMed=2014802;
 RA Hofman K.J., Steel G., Kazanian H.H., Valle D.;
 RT "Phenylketonuria in U.S. blacks: molecular analysis of the
 RT phenylalanine hydroxylase gene.";
 RL Am. J. Hum. Genet. 48:791-798 (1991).
 RN [16]
 RP VARIANTS TRP-252 AND LEU-281.
 RX MEDLINE=91169521; PubMed=1672294;
 RA Okano Y., Wang T., Eisenmith R.C., Longhi R., Riva E., Giovannini M.,
 RA Cerone R., Romano C., Woo S.L.C.;
 RT "Phenylketonuria missense mutations in the Mediterranean.";
 RL Genomics 9:96-103 (1991).
 RN [17]
 RP VARIANT LEU-281.
 RX MEDLINE=91169506; PubMed=1672290;
 RA Dworniczak B., Grudka K., Stumper J., Bartholome K.,
 RA Aulehla-Scholz C., Horst J.;
 RT "Phenylalanine hydroxylase gene: novel missense mutation in exon 7
 RT causing severe phenylketonuria.";
 RL Genomics 9:193-199 (1991).
 RN [18]
 RP VARIANTS SER-48 AND GLY-221.
 RX MEDLINE=91348682; PubMed=1679030;
 RA Konecki D.S., Schlottner M., Trefz F.K., Lichter-Konecki U.;
 RT "The identification of two missense mutations at the PAH gene locus
 RT in a Turkish patient with phenylketonuria.";
 RL Hum. Genet. 87:389-393 (1991).
 RN [19]
 RP VARIANT PKU ILE-94 DEL.
 RX MEDLINE=91236693; PubMed=1709636;
 RA Callaud C., Lyonnet S., Rey F., Melle D., Frebourg T., Berthelon M.,
 RA Vilarinho L., Vaz Osorio R., Rey J., Munnich A.;
 RT "A 3-base pair in-frame deletion of the phenylalanine hydroxylase
 RT gene results in a kinetic variant of phenylketonuria.";
 RL J. Biol. Chem. 266:9351-9354 (1991).
 RN [20]
 RP VARIANTS NON-PKU HPA VAL-306 AND ASN-415.
 RX MEDLINE=93052278; PubMed=1358789;
 RA Economou-Petersen E., Henriksen K.F., Guldberg P., Guettler F.;
 RT "Molecular basis for nonphenylketonuria hyperphenylalaninemia.";
 RL Genomics 14:1-5 (1992).
 RN [21]
 RP VARIANTS PKU GLN-408 AND TRP-408.
 RX MEDLINE=92380641; PubMed=1355066;

RA Lin C.H., Hsiao K.J., Tsai T.F., Chao H.K., Su T.S.;
 RT "Identification of a missense phenylketonuria mutation at codon 408
 RT in Chinese.";
 RL Hum. Genet. 89:593-596 (1992).
 RN [22]
 RP VARIANT 364-LEU-368 DEL.
 RX MEDLINE=93258322; PubMed=1363837;
 RA Jaruzelska J., Melle D., Matuszak R., Borski K., Munnich A.;
 RT "A new 15 bp deletion in exon 11 of the phenylalanine hydroxylase
 RT gene in phenylketonuria.";
 RL Hum. Mol. Genet. 1:763-764 (1992).
 RN [23]
 RP VARIANT LEU-244.
 RX MEDLINE=93258323; PubMed=1363838;
 RA Desviat L.R., Perez B., Ugarte M.;
 RT "A new PKU mutation associated with haplotype 12.";
 RL Hum. Mol. Genet. 1:765-766 (1992).
 RN [24]
 RP VARIANTS PKU.
 RX MEDLINE=94010878; PubMed=8406445;
 RA Guldberg P., Henriksen K.F., Guettler F.;
 RT "Molecular analysis of phenylketonuria in Denmark: 99% of the
 RT mutations detected by denaturing gradient gel electrophoresis.";
 RL Genomics 17:141-146 (1993).
 RN [25]
 RP VARIANT GLY-390.
 RX MEDLINE=93258345; PubMed=8098245;
 RA Abadie V., Jaruzelska J., Lyonnet S., Millasseau P., Berthelon M.,
 RA Rey F., Munnich A., Rey J.;
 RT "Illegitimate transcription of the phenylalanine hydroxylase gene in
 RT lymphocytes for identification of mutations in phenylketonuria.";
 RL Hum. Mol. Genet. 2:31-34 (1993).
 RN [26]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [27]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [28]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [29]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [30]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [31]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [32]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [33]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [34]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [35]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [36]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [37]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [38]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [39]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [40]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [41]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [42]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [43]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [44]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [45]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [46]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [47]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [48]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [49]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [50]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [51]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [52]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [53]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [54]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [55]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [56]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [57]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [58]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [59]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [60]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [61]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [62]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [63]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [64]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [65]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [66]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [67]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [68]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [69]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [70]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [71]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [72]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [73]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [74]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [75]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [76]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [77]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [78]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [79]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [80]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [81]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [82]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [83]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [84]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [85]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [86]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [87]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [88]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [89]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [90]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [91]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [92]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [93]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [94]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [95]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [96]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [97]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [98]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [99]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;
 RN [100]
 RP VARIANT VAL-276.
 RX MEDLINE=94172999; PubMed=8068076;

Query Match 11.3%; Score 213.5; DB 1; Length 452;
 Best Local Similarity 26.6%; Pred. No. 1.4e-09;
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;
 QY 104 RNLVRLSSRFSLWKSYPFRFLDYLEAPGLSLDFLDH-----QAVIKFELETHF 155
 DB 184 KKTGTVFKTLKSLYKTHA---CYEYNHIFPLEKYCGFHEDNIPQLEDVSOFLQCTGF 240
 QY 156 SYIPVSGFVAPHQVLSLLQDRVPPIASVNMRTLDKDNFSLTDLIHLHGHVPLHPSFS 215
 DB 241 RLRPVAGLLSSRDFLGGAPRVFCHCTQYIRHGSKPMYTPEDDICHLLGHVPLFSRSEA 300
 QY 216 EFFTMMGRLFTKVIKVOALPSKKQRIOTLQSNLIAIVRCFWFTVSGLIENHEGRKAYG 275
 DB 301 QFSQEIIG-----LASLGAPDEVIEKLAi-----IYFTVFGCKQGSIKAYG 344
 QY 276 AVLTSSQELCHAFIDNVRLVLELDIIRLPNTSTPOETLFSIRHFDELVE 328
 DB 345 AGLSSFGELQYCLSERPKLLPLELEKTAIQNTVTPEPQPLYVAESFNDAKE 397
 RESULT 21
 PH4H MOUSE
 ID PH4H MOUSE STANDARD; PRT; 453 AA.
 AC P16331;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
 DE monooxygenase).
 GN PAH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_taxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=90241147; PubMed=2334400;
 RA Ledley F.D.; Genet H.E.; Dundar B.S.; Moo S.L.C.;
 RT "Mouse phenylalanine hydroxylase. Homology and divergence from human
 phenylalanine hydroxylase.";
 RL Biochem. J. 267:399-406(1990).
 RN [2]
 RP SEQUENCE OF 12-21.
 RX MEDLINE=80220293; PubMed=7387651;
 RA Weiborn M.; Humble E.; Ragnasson U.; Engstrom L.;
 RT "Amino acid sequence at the phosphorylated site of rat liver
 phenylalanine hydroxylase and phosphorylation of a corresponding
 synthetic peptide.";
 RL Biochem. Biophys. Res. Commun. 93:403-408(1980).
 RN [3]
 RP SEQUENCE OF 277-294.
 RX MEDLINE=85122617; PubMed=6098294;
 RA Robson K.J.H., Beattie W., James J., Cotton R.C.H., Morgan F.J.,
 RA Moo S.L.C.;
 RT "Sequence comparison of rat liver phenylalanine hydroxylase and its
 cDNA clones.";
 RL Biochemistry 23:5671-5675(1984).
 RN [4]
 RP VARIANTS PAH-ENU1 ALA-106 AND PAH-ENU2 SER-263.
 RX MEDLINE=9224471; PubMed=9119379;
 RA McDonald J.D.; Charlton C.K.;
 RT "Characterization of mutations at the mouse phenylalanine hydroxylase
 locus.";
 RL Genomics 39:402-405(1997).
 RL -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
 L-tyrosine + dihydrobiopterin + H(2)O.
 CC -1- COFACTOR: FERROUS ION.
 CC -1- ENZYME REGULATION: N-TERMINAL REGION OF PAH IS THOUGHT TO CONTAIN
 ALLOSTERIC BINDING SITES FOR PHENYLALANINE AND TO CONSTITUTE AN
 "INHIBITORY" DOMAIN THAT REGULATES THE ACTIVITY OF A CATALYTIC
 DOMAIN IN THE C-TERMINAL PORTION OF THE MOLECULE.
 CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- DISEASE: MOUSE STRAINS DEFICIENT IN PHENYLALANINE HYDROXYLASE
 (PAH) WERE CREATED AS MODELS OF PHENYLKETONURIA (PKU).
 CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
 HYDROXYLASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X51942; CAA36205.1; -
 DR PIR; S15758; S15758.
 DR HSP; P00439; 2PAH.
 DR MGD; MGI:97473; Pah.
 DR InterPro; IPR002912; ACT.
 DR InterPro; IPR001273; Aaa_hydroxylase.
 DR Pfam; PF00351; bioplerin_H; 1.
 DR Pfam; PF01842; ACT; 1.
 DR PRINTS; PR00372; FYMHDXKLSE.
 DR TIGRFAMs; TIGR01268; Phehydrox letr; 1.
 DR PROSITE; PS00367; BIOPTERIN_HYDROXYL; 1.
 DR Oxioreductase; Monooxygenase; Phosphorylation;
 KM Phenylalanine catabolism; Iron; Disease mutation;
 FT MOD_RES 16 16 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
 FT METAL 285 285 IRON (BY SIMILARITY).
 FT METAL 290 290 IRON (BY SIMILARITY).
 FT METAL 330 330 IRON (BY SIMILARITY).
 FT METAL 106 106 V -> A (IN PAH-ENU1; MILD PKU PHENOTYPE).
 FT VARIANT 263 263 P -> S (IN PAH-ENU2; SEVERE PKU
 PHENOTYPE).
 FT SEQUENCE 453 AA; 51928 MW; FAICDADDA5968D9D CRC64;

Query Match 11.24; Score 212.5; DB 1; Length 453;
 Best Local Similarity 29.98; Pred. No. 1.7e-09;
 Matches 61; Conservative 35; Mismatches 87; Indels 21; Gaps 5;
 Db 104 RNLM---YRLSSRPSLWVSCYCPPEFLDYLEAF-GLLSDFLDH-QAVIKFELETHFSYX 158
 Db 184 RKTWGVFPFTLKALYKTHACVCHNNHIFPELEKCYGFRFEDNIPLQEDVSQFLQTCGFRUR 243
 QY 159 PVSGFVAPQVYSLLDQRYFFPIASVYRTLDKDNFSLTPDDIHDLGHVPMLLHPSSEFF 218
 Db 244 PVAGLLSDRDFLGLAFVRFHCTQYIRGSKPMYTPEDPICHLLGHVPLFSDRSPAGFS 303
 QY 219 INMGRLFTVIERVOALPSKQRIQTQSNLAIIVRCFMTVSGLIENHGRKAYAVL 278
 Db 304 QEIG-----LASLAPDEVIEKLAT-----IYFTVFGLCRKGDSIKAYAGL 347
 QY 279 ISSPOLGHAFIDNRYVPLPLEDDQ 302
 Db 348 LSSFGELQYCLSDKXKLPLELEK 371
 RESULT 22
 ID PAH_RALSO STANDARD; PRT; 313 AA.
 AC Q8XUJ9;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
 DE monooxygenase).
 GN PHA OR RSC3355 OR RS02630.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxId=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1100;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salamoubat M.; Gatin S.; Artiguenave F.; Gouzy J.; Mangenot S.;
 RA Aljat M.; Billault A.; Brothier P.; Camus J.C.; Catolico L.;
 RA Chandler M.; Choise N.; Claudel-Renaud C.; Cunac S.; Demange N.;
 RA Gaspier P.; Lavie M.; Moisan A.; Robert C.; Saurin W.; Schiek T.;
 RA Weissenbach J.; Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RU Nature 415:497-502(2002).
 CC -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
 L-tyrosine + dihydrobiopterin + H(2)O.
 CC -1- COFACTOR: Binds 1 ferrous ion (by similarity).
 CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
 CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
 HYDROXYLASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL646074; CAD17143.1; -
 DR InterPro; IPR001273; Aaa_hydroxylase.
 DR Pfam; PF00351; bioplerin_H; 1.
 DR TIGRFAMs; TIGR01267; Phehydrox mono; 1.
 DR PROSITE; PS00367; BIOPTERIN_HYDROXYL; 1.
 KM Oxioreductase; Monooxygenase; Phenylalanine catabolism; Iron;
 FT METAL 154 154 IRON (POTENTIAL).
 FT METAL 159 159 IRON (POTENTIAL).
 FT SEQUENCE 313 AA; 34901 MW; 2EBA9E216476371B CRC64;

OR	ENBL; M55915; ARA231115.1; ALT_FRAME.	
OR	ENBL; AF145711; AAD37774.1; -	
OR	PIR; A40986; A40986.	
OR	HSP; P00439; 4PAM.	
OR	Interfero; IPR001273; Aaa_hydroxylase.	
OR	PIR; P00351; bioprotein_H; 1.	
OR	PRINTS; PR00372; FWHYDRXLAASE.	
OR	TIGRFSMS; TIGR01267; Phe4hydrox_mono; 1.	
OR	PROSITE; PS00367; BIOPTRIN_HYDROXIL; 1.	
OR	Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Copper.	
OR	METAL 138 138 COPPER (POTENTIAL).	
OR	METAL 143 143 COPPER (POTENTIAL).	
OR	SEQUENCE 297 AA; 33594 MW; AAA67BG097171FB0 CRC64;	
OR	Query Match 10.4%; Score 196.5; DB 1; Length 297;	
OR	Best Local Similarity 27.4%; Pred.No.1.8e-08;	
OR	Matches 63; Conservative 40; Mismatches 94; Indels 33; Gaps	
QY	126 FLDYLEAFGLSDFL-DHOAVIKFPELETHFSYYPVSGFVAPHOYLSLLQDRYFPFIASVM 184	
DB	63 FLEGLERLEVADRVDPFNKLEKLMATGWKIVAVPGLIPDVVFEEHLNRRFPVTWML 122	
QY	185 RTLDKDNFSLPDLIDLHLGHVPMLLHPSFSEFFINMGRLPFTKVIKQVQALPSSKKQRIOT 244	
DB	123 REPHQLDYLQEPDFHDLFGHVPLLINPVFADYLEAYGKGVKA-KALGALP----- 173	
QY	245 LQSNLIAIVRCFWTFVESGLIENHGRKAYCAVLISSPQE---LGHAFITDNVRVLPLEL 300	
DB	174 -----MLARLYWYTFEGLINTPAGMIYVAGILSKSESIYCLDSASPNRV---GPDFL 224	
QY	301 DQIIRLPNTSQEITLSIRHDELVELTSKLEWMLDGLLESIPLYNQ 350	
DB	225 MRIMWTRITDITFORTYFVDSFKQLFDATP-----DFAPLYLQ 264	
RESULT 24		
OR	PIR; P17276; Q27599; Q27600; O46110; PRT; 452 AA.	
OR	01-AUG-1990 (Rel. 15, Created)	
OR	16-OCT-2001 (Rel. 40, Last sequence update)	
OR	15-JUN-2002 (Rel. 41, Last annotation update)	
OR	Protein henna [includes: Phenylalanine-4-hydroxylase (EC 1.14.16.1)	
OR	(PAH) (Phe-4-monooxygenase); Tryptophan 5-monooxygenase (EC 1.14.16.4)	
OR	(TRH) (Tryptophan 5-hydroxylase)].	
OR	HN OR TPH OR PAH OR CG7399.	
OR	Drosophila melanogaster (Fruit fly).	
OR	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;	
OR	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
OR	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
OR	NCBI_TaxID=7227;	
OR	[1]	
OR	SEQUENCE FROM N.A.	
OR	STRAIN=Canton-S. and Oregon-R; TISSUE=Embryo, and Head;	
OR	MEDLINE=92156168; PubMed=1371286;	
OR	Nackmayner W.S.; White K.	
OR	"A single locus encodes both phenylalanine hydroxylase and tryptophan	
OR	hydroxylase activities in Drosophila."	
OR	J Biol. Chem. 267:4159-4206 (1992).	
OR	[2]	
OR	SEQUENCE FROM N.A.	
OR	MEDLINE=91033030; PubMed=2121612;	
OR	Morales G., Requena J.M., Jimenez-Ruiz A., Lopez M.C., Ugarte M.,	
OR	Alonso C.	
OR	"Sequence and expression of the Drosophila phenylalanine hydroxylase	
OR	mRNA."	
OR	Gene 93:213-219 (1990).	
OR	[3]	
OR	SEQUENCE FROM N.A.	
OR	STRAIN=Canton-S;	
OR	MEDLINE=96332435; PubMed=8769124;	
OR	Ruiz-Vasquez P., Moulard M., Silva F.J.;	
OR	"Structure of the phenylalanine hydroxylase gene in Drosophila	


```

Db 347 LSSVGELEYCLTDKPKLDFE-----PEVTGV---TKYPIOTQFQPLYYVADSFETAKE 396
QY 339 OGL--LESIP-----LYNQKYLSGFEVL 360
Db 397 KTIKANSIPRPGVRYN--AYTQSVEVL 423

RESULT 25
PH4H CAUCR STANDARD; PRT; 294 AA.
ID PH4H CAUCR
AC Q9A7V7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAH) (Phe-4-
monooxygenase).
GN PH4H OR CC1612.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OX Caulobacter
ON NCBI_TaxID=155892;
RP SEQUENCE FROM N.A. / CB15;
RC STRAIN=ATCC 19089 / CB15;
RA MEDLINE=21173699; PubMed=11259647;
RA Nierman W.C.; Feldblyum T.V.; Laub M.T.; Paulsen I.T.; Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
L-tyrosine + dihydrobiopterin + H(2)O.
CC -!- COPACATOR: Binds 1 ferrous ion (BY SIMILARITY).
CC -!- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB005936; AAK23591.1; -.
CC HSSP; P04176; 1PHZ.
CC TIGR; CC1612; -.
CC InterPro; IPR001273; Aaa hydroxylase.
CC Pfam; PF00351; bioprotein H; 1.
CC PRINTS; PR00372; FYWYHDXLASE.
CC TIGRFAMs; TIGR01267; Phe4hydrox mono; 1.
CC PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
CC Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron;
KW Complete proteome.
FT METAL 129 129 IRON (POTENTIAL).
FT METAL 134 134 IRON (POTENTIAL).
FT METAL 134 134 IRON (POTENTIAL).
SQ SEQUENCE 294 AA; 32761 MW; A623276FD0506720 CRC64;

Query Match 10.0%; Score 189.5; DB 1; Length 294;
Best Local Similarity 24.1%; Pred. No. 6.4e-08;
Matches 63; Conservative 43; Mismatches 108; Indels 47; Gaps 6;

QY 118 MKSYCPREFLDYEAFLGSLDPLHQAQVKKFE----LETHFSYYP-----159
DB 24 WETVQAEHDVWITLYPERQDMLHGRACDEFWRGDAALDHLGRGIPDFARINBELKRLTG 83
QY 160 -----VSGFNAPHQYLSLLQDRYFFPTASVMTLDRKDNFSLPDLIHDLLGHVPWLLHPSF 214

```

```

Db 84 WTVAVFGLVPDDVDFDHLANRRFPAGQFIRKPEHLDYLOBPDIHFHDFGHVPMLTDPVF 143
QY 215 SEFFINMGRUFTKVIKQALPSKKORIQTLOSNIILAIKRCFWFTVESGLIENHEGRKAY 274
Db 144 ADY-----MQAYEGGGRALGL-GRLANLARIYWTVEFGLMNTPAGLRIY 188
QY 275 GAVLISSPQELGHAFID-NVRVLPLDQIIRLPENTSTPQETLFSIRHFEDELVELT---330
Db 189 GAGIVSRTSEIFALDDPSNRIGFDLVRVMTLYRIDFQQVYFVIDSIQTLOEVTLRD 248
QY 331 -----SKLEWMLDQGLLEIP 346
Db 249 FGAIYERLASVSDIGVAEIVP 269

RESULT 26
TY3H CAEL STANDARD; PRT; 524 AA.
ID TY3H CAEL
AC P09366;
DT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine 3-monooxygenase (EC 1.14.16.2) (Tyrosine 3-hydroxylase) (TH).
GN CAT-2 OR B0432.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Henkhaus J., Wohlmann P.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC [2]
CC REVISIONS
CC Waterston R.;
CC Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: L-tyrosine, tetrahydropteridine + O(2) = 3,4-
CC dihydroxy-phenylalanine + dihydropteridine + H(2)O.
CC -!- COPACATOR: FERROUS ION.
CC -!- PATHWAY: Catecholamine biosynthesis; first step.
CC -!- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U80836; AAB37888.2; -.
CC HSSP; P04177; ITOH.
CC Wozniak; B0432.5; CE29946.
CC InterPro; IPR001273; Aaa hydroxylase.
CC Pfam; PF00351; bioprotein H; 1.
CC PRINTS; PR00372; FYWYHDXLASE.
CC PROSITE; PS00367; BIOPTERIN HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
KW Neurotransmitter biosynthesis.
FT METAL 352 352 IRON (BY SIMILARITY).
FT METAL 357 357 IRON (BY SIMILARITY).
FT METAL 397 397 IRON (BY SIMILARITY).
SQ SEQUENCE 524 AA; 59593 MW; 32B106950A629802 CRC64;

Query Match 9.4%; Score 178.5; DB 1; Length 524;
Best Local Similarity 24.3%; Pred. No. 9.8e-07;
Matches 77; Conservative 41; Mismatches 114; Indels 85; Gaps 11;

QY 19 KLRQSLSLFFQNSQRAYSTPYRIILQENKEQA-LARH-----KCISILE 69
Db 152 QLHSAELLTQNHVAL-----TKFSIFAKKLSDKNQSQIWFPRHISELQCSKCIKYE 206

```

```

OY 70 FFKLLPVLHLSLKNQRCSTDMNVSTPFENNLYWLLSSRSFSLMKSYPRLFDY 129
D 207 -----PTTDPRHGHGDVAYIARRKFLND-----QALEFGEDEIGY-----VDY 246
OY 130 LE-----AFGLSDP-LDHOAVI-----KPELETH 154
D 247 TEENHATWKAIVYEKLDHLHSHTCAVYRONKILLOEKVLTADRIPOIRDNKFLQKXTG 306
OY 155 FSYVPSGFAVPHOYLSLQDYPFASVMTRLDNDNSLRDILHDLGHVPMILHPSF 214
D 307 FELRFGSLISARDFLASLAFVFTTTLRHHSKPHHSPEPDLHELGHVPMFSDPL 366
OY 215 SEFFINNGRLFTVIEKVALPSKKORIGTLOSNIATVRCGFTVSSGLIENHGRKAY 274
D 367 AQMSQDIG-----LMSLGASDEHIEKLSF-----VWFIVEFGLCEDGKLKAI 410
OY 275 GAVLISSPQELGHAFID 291
D 411 GAGLLSAYGELMHACSD 427

RESULT 27
PH4H CAEEL STANDARD; PRT; 457 AA.
AC P90925;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative phenylalanine-4-hydroxylase (EC 1.14.16.1) (PAM) (Phe-4-monooxygenase).
GN K08F8.4
OS Caenorhabditis elegans.
OC Bursarctica, Metazoa, Nematoda, Chromadorea, Rhabditida, Rhabditioidea,
OC Rhabditidae, Rhabditinae, Caenorhabditis.
OX K08I_taxid=6239;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RL Smye R.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-phenylalanine + tetrahydrobiopterin + O(2) =
CC L-tyrosine + dihydrobiopterin + H(2)O.
CC -1- COFACTOR: FERROUS ION (BY SIMILARITY).
CC -1- PATHWAY: Catabolism of phenylalanine; first (rate-limiting) step.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z66497; CAA31286.1; -
DR HSBP; P04176; 1FHZ.
DR Wormpep; K08F8.4; CE21050.
DR InterPro; IPR002912; ACT.
DR InterPro; IPR001273; Aaa_hydroxylase.
DR Pfam; PF00351; Bioplerin_H; 1.
DR Pfam; PF01842; ACT; 1.
DR PRINTS; PR00372; FYMHYDRKLASE.
DR TIGRFAMs; TIGR01268; Phehydrox_lectr; 1.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KM Hypothetical protein; Oxidoreductase; Monooxygenase;
KM Phenylalanine catabolism; Iron.
FT METAL 285 285 IRON (BY SIMILARITY).
FT METAL 290 290 IRON (BY SIMILARITY).
FT METAL 330 330 IRON (BY SIMILARITY).
SQ SEQUENCE 457 AA; 52129 MW; 68365836DFEC8DAF CRC64;

```

```

Query Match 9.3%, Score 175.5; DB 1; Length 457;
Best Local Similarity 25.5%, Pred. No. 1.4e-06;
Matches 56; Conservative 35; Mismatches 80; Indels 49; Gaps 5;

OY 143 QAVIKFPELETHFSYVPSGFAVPHOYLSLQDYPFASVMTRLDNDNSLRDILHDL 202
D 228 QDVSDFLKDCTGTITRPAGLSSNDPLAGLAFVHSITQYIHHSAPKYTEPDI CHEL 287
OY 203 LGHVPMLHPSRSPERFIMGRFLF-----TKYIEKVALPSKKORIGTLOSNIATVRCMF 258
D 288 LGHVPFLFAVVERKQTSQIEGLASLQADPDVIEKATL-----YWF 327
OY 259 TVESGLIENHGRKAYGAVLISPOELGHAFIDNVRLPLE--LDQITLPRNTSPQET 316
D 328 TIEFGICQDDGKKAVGAGLSSFGELQYALSDPEVDDPRAVCCVTIKPTTEYQPKTF 387
OY 317 L-----FSIRH--FDELVELTSKL 333
D 388 LAESFASAKKUKLSMAATINRPFQIRYNAATYORVELDKV 427

RESULT 28
NINQ DROME STANDARD; PRT; 1501 AA.
AC P10676; P10677;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neither inactivation nor afterpotential protein C (EC 2.7.1.37).
GN NINAC.
OS Drosophila melanogaster (fruit fly).
OC Insecta, Arthropoda, Mandibulata, Pancrustacea; Hexapoda;
OC Insecta, Preygora, Neoptera, Endopterygota, Diptera, Brachycera;
OC Muscomorpha, Phyllophaga, Drosophilidae, Drosophila.
OX NBI1_taxid=7227;
RA [1]
RP SEQUENCE FROM N.A., FUNCTION, AND ALTERNATIVE SPLICING.
RX MEDLINE=88151067; PubMed=2449973;
RA Montell C.; Rubin G.M.;
RT "The Drosophila ninaC locus encodes two photoreceptor cell specific
RT proteins with domains homologous to protein kinases and the myosin
RT heavy chain head."
RL Cell 52:757-772 (1988).
CC -1- FUNCTION: REQUIRED FOR PHOTORECEPTOR CELL FUNCTION. THE NINAC
CC PROTEINS JOIN PUTATIVE SERINE/THREONINE-PROTEIN KINASE AND MYOSIN
CC ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
CC short form; are produced by alternative splicing.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE SER/THR
CC FAMILY OF PROTEIN KINASES.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO THE MYOSIN
CC SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J03131; AAA28718.1; -
DR EMBL; J03131; AAA28719.1; -
DR EMBL; M20230; AAA28721.1; -
DR EMBL; M20231; AAA28720.1; -
DR PIR; A29813; A29813.
DR PIR; B29813; B29813.
DR HSBP; P08799; 1MND.
DR FlyBase; FBgn002938; ninaC.
DR InterPro; IPR000719; Euk_kinase.

```



```

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21648401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Collins M., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellinelli T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby I., Howarth S., Huckle B.J., Hunt S., Jagsis K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Pearson D., Quail M.A., Rabinowitch E.,
RA Oliver K., O'Neill S., Saunders D., Murphy L., Niblett D., Ocell C.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymopre B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Drees S., Gloux S., Leclaire V., Mottier S.,
RA Gallier F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lacas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forstburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Spakowski G.V., Usery D., Bartoli B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable)
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 269086; CA93165.1; -
CC HSSP; P12351; 1PYC.
CC InterPro; IPR001138; Fungi_Trn.
CC Pfam; PF001172; Zn_clus; 1.
CC SMART; SM00066; GAL4; 1.
CC PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
CC PROSITE; PS00463; ZN2_Cy6_FUNGAL_2; 1.
CC Hypothesis: protein; Transcription regulation; DNA-binding;
CC Nuclear protein; zinc; Metal-binding;
CC DNA BIND 19 45
CC SEQUENCE 563 AA, 64690 MW, C32CDBRAFA6dD38 CRC64;
SO
Query Match 5.4%; Score 102; DB 1; Length 563;
Best Local Similarity 21.1%; Pred. No. 1.1;
Matches 80; Conservative 56; Mismatches 142; Indels 102; Gaps 16;
Db
7 TLDPKYLKIALKRLQSLSPFONSQSLORAVSTPYSYRITLLOKENEKOLARRKCS 66
185 TLPIIAATITQSLDPLDVLINFNAGI-----TPLESSRIINKAKNEISEOEYKHICLP 239
Qy ILLEFFKULLFVHLISLKNQREGCSTDM-----AVSTPEFNNILVYRLSLRSFSLWKS 120
Db 240 DKEIITQWLLRAVATKFRTRIRGVNTDLCSRIHSTVLPPLF-QVTEKICKNTSLMWA 297
Qy 121 YCRFPIDYLEAFGLSDPLDHOAVIKFELETHPSYVPGVAFAPHOYVLSLQDRYFPI 180
Db 298 LC-----EIDGL-----ECVLKY--RPFIDHDTYGRLLK----- 325

```


CC -----
 DR EMBL; U39688; AAC71293.1; -
 DR EMBL; U01715; AAC43189.1; ALT_INTT.
 DR EMBL; U02251; AAD12514.1; -
 DR EMBL; U01749; AAD10562.1; -
 DR EMBL; U01775; AAD10595.1; -
 DR TIGR; MG075; -
 FM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 11 31 POTENTIAL.
 FT SEQUENCE 951 971
 SQ SEQUENCE 1024 AA; 116424 MW; 5B15406855CB554 CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 1024;
 Best Local Similarity 19.9%; Pred. No. 4.6;
 Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

DB 29 QNSQSLQAVSTPYSYRILOKENEK---QALARRKICISILEFFKYLFPVHLSISK 85
 DB 564 QQTSLKSLFSV---IGDILSETVNRKTTLHAAVKNNELSLIVETASTLKIKHL----- 613
 DB 86 QREGCTDMAVSTPFRNRLMYRLSRSFSLMKSCYCRFFLDYLEAFGLSDPLDQAV 145
 DB 614 -----NVQYKVLVDKFEIKNSFIK-----ELINFPDQKIDTP 647
 DB 146 IK--FPELETH-----FSYVPGFVAPRHOYSLIQDRYFPPLASVMTLDKDNF 192
 DB 648 IKKYLFESENYKTLRKKEENEGFPGYMAKFTVPGTFSANTYSAL-----DKT 686
 DB 193 SLTPDLTHDLGHVPLLRPSSEFPIINMGRLFT-----KYLEKVALPSKKRIQT 244
 DB 699 KSIIDLADNHL-FGKSLESVNDSDFIKINSFLLKYGDNLNLPVHSLITKQVGYOI 757
 DB 245 LQSLALAIYRCFWFTVESGLIENHGRKAYGAVLISPOELG----- 286
 DB 758 VNVVFHIDALLTAELOAVFSNPK-----FVKSFEVLSKSLFEVWKTIPENSVOI 810
 DB 287 ---HAFIDNVRVLPLELDQIIRLPNNTSPOETLFSIRHFD 324
 DB 811 LKKEYTEKDNLKFFPKADGSSRLPFDLSKPDQAVIPPAFVD 852

RESULT 34

CP1 PANAR STANDARD; PRT; 492 AA.

AC Q27712;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 2L1 (EC 1.14.14.1) (CYP2L1).
 GN CYP2L1.
 OS Panulirus argus (spiny lobster).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 OC Malacostraca; Decapoda; Eucarida; Decapoda; Pleocyemata;
 OC Palinura; Palinuroidea; Palinuridae; Panulirus.
 OX NCBI_TaxID=6737;

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC MEDLINE=96201120; PubMed=8619632;

RA Jamee M.O., Boyle S.W., Ripido-Rosenthal H.G., Smith W.C.,

RT "CDNA and protein sequence of a major form of P450, CYP2L1, in the

RT hepatopancreas of the spiny lobster, Panulirus argus."

RT Arch. Biochem. Biophys. 329:31-38(1996).

CC - FUNCTION: EFFICIENT IN CATALYZING THE MONOOXYGENATION OF

CC BENZHEXYLAMINE, AMINOPYRINE, BENZO (A) PYRENE, PROGESTERONE, AND

CC TESTOSTERONE.

CC - CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +

CC oxidized flavoprotein + H(2)O.

CC - SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

CC similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U44826; AAB03106.1; -
 DR HSSP; P00179; 1DT6.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum.
 FT BINDING 436 436 HEME (BY SIMILARITY).
 FT SEQUENCE 492 AA; 56767 MW; F27EE702D09D5EE CRC64;

Query Match 5.1%; Score 97; DB 1; Length 492;
 Best Local Similarity 19.9%; Pred. No. 2.4;
 Matches 74; Conservative 56; Mismatches 106; Indels 136; Gaps 18;

DB 56 KQALARRKICISILEFFKYLFPVHLSISKORREGCTDMAVSTPFRNRLMYRLSSRF 115
 DB 79 KTLKSRFCSDDPDTYTKLF-----GENDVGV---FSNGV----- 113
 DB 116 SLMKSCYCRFFLDYLEAFGL-----LSDFLDQAVIKFELETHFSY---YFVS----- 161
 DB 114 -MMQTH-KRFTLRQLDLMGKSRLEAIOEAAACLVQELKRTDQMPPLPSINLAVLN 171
 DB 162 ---GFVAPRHOYSLIQD--RFPPLASVMTLDKDNFSLTPDLTHDLGHVPL--HPS 213
 DB 172 VIKKLVADHRY--SLDDEGQFTQLLTITDNNQGFAL-----NLFFYLWLLMTTPD 223
 DB 214 FSEFPIINMGRLFTKYLEKVALPSKKRIQTLSNLAIYRCFWFTVESGLIENHGRKA 273
 DB 224 FVKNMGVAVLRDGVCE-----LKDYMKT-----FKEHQ----- 253
 DB 274 YGAVLISPOELGHAFF-----IDNVRVLP----- 298
 DB 254 -ATLDPSNPKDLIDAVLIDLOERKEDPLSTNNITETVAVIMDFGAGTETSTMIRWTL 312
 DB 299 -----ELDQIIRLPNNTSPOETLFSIRHDELVELTSKLEWMLDQGLLESIP----- 347
 DB 313 YLKKYEVQAKIQREIDAAVRGTLPSEHKDKLAYEATIHVV--HRIVSLVPLGVSHY 370
 DB 348 YNOKYLSGFV 359
 DB 371 TNQDTLQGYRL 382

RESULT 35

EX58 CHLPN STANDARD; PRT; 1050 AA.

ID 092767; G93581; G932F2.

DT 30-MAY-2000 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Exodeoxyribonuclease V beta chain (EC 3.1.11.5).

GN RCB OR CPN0738 OR CPN007.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.

CC NCBI_TaxID=83558;

CC (1)

CC SEQUENCE FROM N.A.

CC STRAIN=CML029;

CC MEDLINE=9920606; PubMed=10192388;

RA Kaiman S., Mitchell W., Marache R., Lammel C., Fan J., Hyman R.W.,
 RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";
 RL Nat. Genet. 21:385-389(1999).

[2]

RP SEQUENCE FROM N.A.
 RX STRAIN=AR39;
 RA MEDLINE=20150255; PubMed=10684935; Gill S.R., Heidelberg J.F.,
 RA Read T.D., Brunham R.C., Shen C., Gyll S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Quinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN=J138;
 RA MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Iehi K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314 (2000).
 CC -!- FUNCTION: REQUIRED FOR EFFICIENT DNA REPAIR; IT CATALYZES THE
 UNWINDING OF DOUBLE-STRANDED DNA AND THE CLEAVAGE OF SINGLE-
 STRANDED DNA AND IT STIMULATES LOCAL GENETIC RECOMBINATION.
 CC ALL OF THESE ACTIVITIES REQUIRE CONCOMITANT HYDROLYSIS OF ATP
 (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage (in the presence of
 ATP) in either 5' to 3' or 3' to 5'-direction to yield 5'-
 phosphooligonucleotides.
 CC -!- SUBUNIT: CONSIST OF THREE SUBUNITS; RECB, RECC AND RECD
 (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. UVDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE001655; AAD19877.1; -;
 DR EMBL; AE002164; AAF37903.1; -;
 DR EMBL; AP002547; BAA98945.1; -;
 DR HSSE; P09980; IUA.
 DR TIGR; CP0007; -;
 DR InterPro: IPR000212; Uvdr-helicase.
 DR Pfam; PF00580; Uvdr-helicase; 1.
 KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
 KW DNA repair; Complete proteome.
 FT NP_BIND 21 28 ATP (POTENTIAL).
 FT CONFLICT 142 142 N -> K (IN REF. 2).
 FT CONFLICT 182 182 V -> I (IN REF. 1).
 SQ SEQUENCE 1050 AA; 121056 MW; 6E2CD03C2AEB83B7 CRC64;
 Query Match 5.1%; Score 97; DB 1; Length 1050;
 Best Local Similarity 22.1%; Pred. No. 6.2;
 Matches 77; Conservative 51; Mismatches 117; Indels 104; Gaps 19;
 QY 46 RIILQKNEKQALAHKICISL--EFFKNLFF---VHLLSLSKNQREGCGTDMA-----95
 DB 137 RLHKNPALPHTSGLVHHTNYLKLQNLKXNLFQEQFHLLAVRYNVTSKTSSLVLDKLLA 196
 QY 96 ----VSTPFENR-----NLWV-RLSSRSLSKSCYPRFFLDYLEA-----FG 134
 DB 197 SYTQPISSVFSFSSRVERLEQSLWHQIQVNSLLEIKP-----QVFLDQLTAHISGFKKQPS 252
 QY 135 LLSDFLDHQAIVKFELETH--FSYYPVS-GFVAPHQVLSLODRYFPITAS-----182
 DB 253 ILDDU--HHFVLLYTSETSSLSFSFKFAETNFKRLA-----RYKCAATVLENWS 305
 QY 183 -VWRTLDKDNF-----SLTDLTDLHDL-LGHVPWLHPSFSSEFFINMRLTKVIEKQAL 235
 DB 306 WVERTLEPCNDLRIFNTLLDQLQYLYKQNTW-LSPDSVF-----ALEKLSS 354

QY 236 PSKORIQTQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPOELGH---AFIDN 292
 DB 355 SEAPVVAQ-----REQYQLVLIDEFQDTDKQWSIFS 389
 QY 293 VRVPLELDIILRPENTSTPOETLFSIRHDELVELTSKLEWMLDOGL 341
 DB 390 LFISPKTKTGL-----FLIGDPKOSIYEWASADLPYLITAKSFSFSEDKQL 434
 RESULT 36
 G6PD_CHLFPN STANDARD; PRT; 512 AA.
 ID G6PD_CHLFPN
 AC Q928U6; Q9J022;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
 GN ZWF OR CPN0238 OR CP0524 OR CPJ0238.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Quinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Iehi K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314 (2000).
 CC -!- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADPH.
 CC 1,5-lactone 6-phosphate + NADPH.
 CC -!- PATHWAY: Pentose phosphate pathway; first step.
 CC -!- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE001609; AAD18391.1; -;
 DR EMBL; AE002211; AAF73682.1; -;
 DR EMBL; AP002545; BAA98448.1; -;
 DR HSSE; P11411; 1DPG.
 DR TIGR; CP0524; -;
 DR InterPro: IPR001282; G6PD.
 DR Pfam; PF00479; G6PD; 1.
 DR Pfam; PF02781; G6PD_C; 1.


```
DE Large tegument protein.
GN U31.
OS Human herpesvirus (type 7 / strain J1) (HHV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=57278;
RN [1]
RP SEQUENCE FROM N.A.
RA Nicholas J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EH-1 24, EBV BPLF1, HSV-1 64, VZV 22, AND HCMV UL48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U43400; AAC54693.1; -
CC SEQUENCE 2059 AA; 239476 MW; AAE6CA6DEDC5D316 CRC64;
SQ
Query Match 5.1%; Score 96; DB 1; Length 2059;
Best Local Similarity 20.6%; Pred. No. 18;
Matches 90; Conservative 62; Mismatches 164; Indels 120; Gaps 23;
QY 13 ILKIALKRLQSLFFQNSQS-LQRAYSTPYGYRIILQKENKEQALARKH--CISILE 69
DB 1265 ILKWLIVPVKELNTFFVATWSEFGEVDPDYKHFR-ALEYELNSKIYEIKNICIEIIE 1323
QY 70 PFENL--LFVHLLSLSKNOREGCGSTD---MAVSTPFENLWYRLLSRF--SLWKSXY 121
DB 1324 NTDNTEKLSLTIKQIDNRIAGGCKQFQDYSKILTAETNQQ-----QTRYEQELKKQY 1377
QY 122 CRRPFLDYLEAFGLLSDFLDHOAVI-----KPELTHSYYP-----VSGF 163
DB 1378 FD--LLDNIAHFRAPDFNQHQNLILKDKDKTKTDTIVFERFNLDDTFVSMVNEFP 1435
QY 164 VAPHOYLS-----LLQRYEPI-ASVMTLDDKDNFSL-----TP- 196
DB 1436 LOALEALSHVQAAQNFQNLVTEQADLPQNFIPVSLSTVTKTIPKSDINLRMKIHTPQ 1495
QY 197 -----DLIHDLLGHVPWLLHPSEFSEFFINMGR-L-FTKVI--EKVQALPSKKQRI 242
DB 1496 TFFQVDSVFNTQLIVDEKG-IPVQFYNNFHNIVKFFPALNYKKIIVPKVLNVLSTKYKI 1554
QY 243 QTLQSNLIAIVRCFW-FTVBSGLIENHEGR-----KAYCAVLIS-----SPQ 283
DB 1555 LITLKSILSVKSKWEIINFDLTSYFQKAEFTFQNVFPFIINLKIFIYIITQAWSVTSD 1614
QY 284 ELGHAFIDNVAVLPLEDQIIRLPENTSTPQETLSIRHFDVELTSLKLEWMDQGLLE 343
DB 1615 ETQHSF-----ELEKFSLLIANN-----PEFLF-----GSLQCPVDLAINS 1653
QY 344 SIPLYNQEKYLSGFEV 359
DB 1654 LIPLLEKKKYFTATI 1669
RESULT 39
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
```

```
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Chessman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RL "The gene encoding topoisomerase II from Plasmodium falciparum.";
Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X79345; -; NOT ANNOTATED_CDS.
DR HSSP; P06786; 18GW.
DR InterPro; IPR003594; ATPbind_Atpase.
DR InterPro; IPR003957; CBFA_NFYB_topis.
DR InterPro; IPR001241; DNA_topoisom.
DR InterPro; IPR002205; DNA_topoisom.
DR Pfam; PF00204; DNA_gyraseB.1.
DR Pfam; PF00521; DNA_topoisom.
DR Pfam; PF02516; HATFase_C.1.
DR PRINTS; PR00615; CCNATSUBUNTA.
DR PRODOM; PR00418; TP12FAMILY.
DR PRODOM; PD000616; DNA_topoisom.1.
DR PRODOM; PD000742; DNA_topoisom.1.
DR SMART; SM00433; TOP2C; 1.
DR SMART; SM00434; TOP4C; 1.
DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88F83BE9 CRC64;
Query Match 5.1%; Score 95.5; DB 1; Length 1398;
Best Local Similarity 20.8%; Pred. No. 12;
Matches 87; Conservative 55; Mismatches 149; Indels 127; Gaps 21;
QY 18 LKLRQSLFPQNSQSLORAYSTPYGYRIILQKENKEQALARKHKSILEPKNL--- 74
DB 601 LKHGFLSEFVTPIVKQKG-SQEYFFIIAEYEQWKNLTNLGK-----IKYKGLGTS 655
QY 75 ---LFVHLLSLSKN-----QREGCTDMNAVSTPFFNRLWYRLLSRFSFLWKSYS 122
DB 656 TDREFKQYFDIKKHKIMFLWTGDRGDSIDMAFSKRIEDRKLW----- 700
QY 123 PRFFLDYLEAFGLLSDFLDHOAVIKFPELETHFSYYPVSGFVAPHQYLSLLQDRYPIAS 182
DB 701 : : : : :
DB 701 -----LQNF-ILGSYVDHK-----EKDLSY---DFVNKE---LIYGRYDTERS 738
QY 183 VMRTLD-----KDNFSLTPDLIHDLLGHVPWLLH---PGFSFEEFINMGR 223
DB 739 IPNIDMGKPGQRKVLGYCFCFKRLNCKEQAQLVGVIAEAGVHHHGESSLQQTIIINMAQ 798
```

```

0Y 224 LE-----KYLEKVALPSKKO-----RIGTOSNIJAIYRCMFWFVSGGLE--NHG 270
Db 799 TFGVSNININLEFCGCGSRKEGCKDASAKRIYFIPLKLSSTRSIFNEYDDPLIKYINERG 858
0Y 271 RK-----ATGAVLISFDELGNF-----INVR-----VPLEL 300
Db 859 QKIEPQYIYIPITLILNCGEGIGTGISFFIPNVYKCIIONIKRYNKEPLIPMPFWYK 918
0Y 301 DQILRLPFTSTPOETLFSIRHFD---ELVELTSKLEWMD-OGLLIESPLYNQEKY 353
Db 919 DPKRIESNKTGYETIGIINKIDNDLTLEITELPK-KWTQDYKFELE--LTTDEKH 973

RESULT 40
SRB8_YEAST
ID SRB8_YEAST STANDARD; PRT; 1427 AA.
AC P25648; P25647;
AD 01-MAY-1992 (Rel. 22, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Suppressor of RNA polymerase B SRB8.
GN SRB8 OR YCR081W OR YCR81W/YCR80W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC BAKARYOTA; FUNGI; ASCOMYCOTA; SACCHAROMYCOTA; SACCHAROMYCETES;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycet.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95293223; PubMed=7774808;
RA Hegartner C.J., Thompson C.M., Zhang J., Chao D.M., Liao S.M.,
RT Koleske A.J., Okamura S., Young R.A.;
RT "Association of an activator with an RNA polymerase II holoenzyme.";
RL Genes Dev. 9:897-910(1995).
RN [2]
RP SEQUENCE OF 1-530 FROM N.A.
RA Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
RL Sanz E.;
RN Submitted (MAR-1992) to the EMBL/genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 531-1427 FROM N.A.
RA Feidmann H., Mannhaupt G., Vetter I.;
RL Submitted (MAR-1992) to the EMBL/genbank/DBJ databases.
RN [4]
RP REVISIONS.
RA Gromada R.;
RL Submitted (JAN-1996) to the EMBL/genbank/DBJ databases.
CC -1- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE
CC MEDIATOR OF ACTIVATION SUBCOMPLEX.
CC -1- SUBCELLULAR LOCATION: Nucleus (Probable).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; X59720; CAA4266.1; -
DR PIR; S19496; S19496
DR PIR; S19495; S19495.
DR TRANSFAC; T02152;
DR SGD; S0000677; SRB8
DR InterPro; IPR000651; RASGEF4
KW Nuclear protein.
KW SEQUENCE 1427 AA, 166859 MW, 776A29AD82E331FE CRC64;

```

```

Query Match      5.0%; Score 95; DB 1; Length 1427;
Best Local Similarity 18.8%; Pred. No. 13;
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY      12 YTLKALTKROSLSIFPNQNSIQRAVSYPPYYII-----LQENKEKQALARKKL 65
|||||  |||  ::  ::  |||  |||  :  |||  :  |||  :  |||  :  |||  :

```

```

Db 214 YILE---KLIIPMTNHYVDSQUL-RTWKQRI5YFLKLGNCVSLRLINKE---IFHHWV 286
Qy 66 SILEPKNLFL---VHLISLSKNGRGCGSTDAVYVST---PPF- 102
Db 267 EPINKMENEELPLSLHILMLFMDICQIDITNAPVAATITTSQKEPFLVTKITDMLLHK 326
Qy 103 -----NRNLWRLLSSRFL-----WKS 120
Db 327, YVIVSSKSMINDENYIINDIKNNKKFKMLIKLISLLIKLIPQOSLEVEIPTSMWEI 366
Qy 121 YCFRPF-----LDYLEAFGLSDPLDQAVIKFPELTFHSYYPGCVAF 166
Db 387 YKFLFEIVSNADTQNSDMMKKKLELISYNESSLKUNSSIR-----NVIMASNNAN 457
Qy 167 HQYLSLDORYPPLASVW-----RTLDKDN-----FSLPDLIHLLGHVP 207
Db 438 DPDLITVTKQKPFKSCQLNCIDTQFTKLD-DNTEPDEWPYTYDQNPFLMHKIIQIL 456
Qy 208 WILHPS--FSEPFIMKGLFPKVIKYOALPSSKQRIQTLQSNL--IAYRCFVFTYESG 263
Db 497 WSHIPSRQPDHYESN-----QLVAKLLL-----RINSTDBLHEFOIEDAIVSLVFO- 544
Qy 264 LIIINHGRK-----AYGAV-----LISSP-----OELIHAFFD-- 291
Db 545 LAKNFSAQKRVVSYVMPSLRYLRLNILITTYGIIKVPYIRKLLISGLLYLQDSNDFVHQ 604
Qy 292 ---NVAVLPLELDOIIRLPEFTSPQETSLSIHHPFELVELTSKY----- 333
Db 605 LILNKISPLMSQYNMVLKRVWMEYDVAKFEIENFQVLVEIEFOIKMILISNDITNLQLS 664
Qy 334 -----EWMLD---QSLSEIPLVNOEKYLSGFEVLC 361
Db 665 KTLPSIKIWAWEVYLSHLCSGSLSSV--NRVLLKIFKIFPC 703

```

```

RESULT 41.
TRA2 CAEEL.
ID ID TRA2 CAEEL. STANDARD; PRT: 1475 AA.
AC p34709;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1999 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sex-determining, transformer protein 2 precursor.
GN TRA-2 OR C15F1.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
CX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=92360913; PubMed=1498366;
RA Kuwamura P.E., Okkema P.G., Kimble J.;
RT "tra-2 encodes a membrane protein and may mediate cell communication
RT in the Caenorhabditis elegans sex determination pathway.";
RL Mol. Biol. Cell 3:461-473(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX Bentley D., Pavello A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.J.
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROMOTES FEMALE DEVELOPMENT IN XX ANIMALS WHERE IT
CC SEQUESTERS ONE OR MORE OF THE FEM PROTEINS TO THE MEMBRANE THEREBY
CC FREEING THE TRA-1 PROTEIN (A PUTATIVE TRANSCRIPTION FACTOR) TO
CC ENTER THE NUCLEUS AND PROMOTE FEMALE DEVELOPMENT. IN XO ANIMALS IT
CC ACTS AS A RECEPTOR FOR HER-1 WHICH PREVENTS IT FROM BINDING TO FEM
CC PROTEINS THEREBY REPRESSING THE ACTIVITY OF TRA-1.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: SOMATIC AND GERM LINE TISSUES.

```


Db 112 EANAATFPVDDAPVXKLPVTEFLKVDNDNFKPLNPKGLNISDNIALPFRVNLKSIEL 191
 QY 190 DNEFSLT-----PDLIHLHLGHVFWLHPSESEPFIMKGLFTKVIKVOALPSSKKQRIQTL 245
 Db 192 KMFDAFLRKADVITIKKQKRYNFMOMQOTYVGGNTVINLDTLQAQVNFANIQDL 251
 QY 246 QSNLIAT-----VRCFWFVSGLIEN-----HEGRKAYGAVLISSEFQELGHAFIDNVR 294
 Db 252 QNTFVKVGNLSTQLEWIFPTVKNKLTQNGNDLTHAKTVIGSEFFQTNVNLAKSVLEIDK 311
 QY 295 VLPL---ELDQITRLPF 308
 Db 312 VQPLVKQAFERVLTPF 328

RESULT 43
 ID _Y065_MYCGE STANDARD; PRT; 466 AA.
 AC P47311: 049281;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical ABC transporter ATP-binding protein MG065.
 GN MG065.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 NC NCB1_TaxId=2097;
 OK NCB1_TaxId=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569933;
 RA Frieser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403 (1995).
 RN [2]
 RP SEQUENCE OF 393-466 FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075220; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing";
 RL J. Bacteriol. 175:7918-7930 (1993).
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39686; AAC71283.1; -
 DR EMBL: U02154; AAD12436.1; -
 DR TIGR: MG065; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_tran.1.
 DR Pfam: PF00005; ABC_tran.1.
 DR ProDom: PD000006; ABC_transporter.1.
 DR SMART: SM00218; AAA_1
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM Hypothetical protein; ATP-binding; Transport; Complete proteome.
 FT NP_BIND 269 276 ATP_BINDING POTENTIAL.
 FT GO 394 399 GEFPTG -> WWITW (IN REF. 2).
 SQ SEQUENCE 466 AA; 54009 MW; A585044B8A90391C CRC64;

Query Match 4.9%; Score 92.5; DB 1; Length 466;
 Best Local Similarity 22.0%; Pred. No. 5;

Matches 61; Conservative 44; Mismatches 95; Indels 77; Gaps 15;
 QY 68 LEPFKLLFVHLISLSTKQREGCSTDAVAVSTPFENRNLMLRLSSRSLMK----- 119
 Db 31 LEEQNYLQKGLKQKKEYSKI-----IYNPLEK---RFKMKENTETFE 76
 QY 120 SYCPFFFLYL- -AFGLSDPLDQAVIKFELETHFSTYPSGVAAHQYLSL----- 172
 Db 77 SYDRFFFTKYNHYSLSLPSFINEQ-----LETVIASY--NSPLNEHNLAKNKSQF 127
 QY 173 -LDRYFPPIASVMTLDDKNFSLTPLD-----IHLGHVFWLHP 213
 Db 128 SFPEKLFELATQGFNNLEK-NTAISDDLPLQKRYRTQLKAQREERELKLNLIK-LKILS 185
 QY 214 -----FSEPFIMKGLFTK-VIEKVALPSKKQRIQTLQ--SNLIAVRCFWFVSS 262
 Db 186 EKQGEILLNNWFSNERLFLKNEVKVWNLNSPRQQAQIDQNIIEIKVVKY-ITN 244
 QY 263 GLIENHGR-----KAYGAVLISSEFQELGHAFIDNVR 293
 Db 245 GITTNVAVGVDLAKSHDFVILGSGSGKTTILNI 281

RESULT 44
 ID _FLP_KLULA STANDARD; PRT; 447 AA.
 AC F13783;
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 01-JUN-1990 (Rel. 13, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Recombinase FLP protein.
 GN A.
 OS Kluyveromyces fragilis (Yeast).
 OG Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Kluyveromyces.
 NC NCB1_TaxId=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86232585; PubMed=3520486;
 RA Chen X.J., Sallio M., Falcone C., Bianchi M.M., Fukuhara H.;
 RT "Sequence organization of the circular plasmid pK01 from the yeast
 RT Kluyveromyces fragilis";
 RL Nucleic Acids Res. 14:4471-4481 (1986).
 CC -1- FUNCTION: FLP CATALYZES RECOMBINATION BETWEEN THE LARGE INVERTED
 CC REPEATS OF THE PLASMID.
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X03961; CAA27591.1; -
 DR EMBL: X03961; CAA27591.1; -
 DR PIR: S28086; S28086.
 KM DNA recombination; DNA integration; Plasmid.
 FT ACT_SITE 336 336 TRANSIENT COVALENT LINKAGE TO DNA DURING
 FT STRAND CLEAVAGE AND REJOINING (BY
 FT SIMILARITY).
 SQ SEQUENCE 447 AA; 51110 MW; 89D08AB52718A104 CRC64;

Query Match 4.9%; Score 92; DB 1; Length 447;
 Best Local Similarity 23.6%; Pred. No. 5.2;
 Matches 54; Conservative 37; Mismatches 82; Indels 56; Gaps 12;
 QY 13 LKIALKRGOSLSPFQNSQSLQRAVSTPSTYRIILQK-ENKEKALAHKICSLTEFF 71
 Db 81 LKFTVRLK-----TESLQKIESALPSTYKVVSPFKNOEVLDFRYE-----ETH 127
 QY 72 K-----NLFLVHLISLKN-----QREGCSTDAVAVSTPFENRNLMLRL-----SS 113

Db 128 KYDSWGLQNTILSKEKIDKIVIRIACFFDQSCVTT---TRAERYLLLLGAVGNCC 184
 QY 114 RSLWKSVCPRFFLDYLEAGLLSDFLDHQAVIKFELETHFSYYPVSG---FVAPHQY 169
 Db 185 RYSDLNLDPTFTFIYNNST--LGPVIRATVTETKSTERYVNFYVNGDCDLLISLYDY 242
 QY 170 LSLQDRYFPIASVMTLTKDNFSLTP-DLIHDLGLHVPWLLHPSFSEF 217
 Db 243 L-----RVCSPIEKTIVSSNRTQTQHFL---PESLARTFSRF 277

RESULT 45

MATK ARATH
 ID MATK ARATH STANDARD; PRT; 526 AA.
 AC P56784;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable intron maturase (Maturase K).
 GN MATK OR YCF14
 OS Arabidopsis thaliana (Mouse-ear cress).
 OG Chloroplast
 OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=cv. Columbia;
 RC MEDLINE=20039611; PubMed=10574454;
 RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.;
 RT "Complete structure of the chloroplast genome of Arabidopsis
 thaliana.";
 RL DNA Res. 6:283-290(1999).
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS.
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS.
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
 CC BY MITOCHONDRIAL INTRONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF000423; BA84366.1; -.
 CC InterPro; IPR000442; Intron_mature2.
 CC InterPro; IPR002866; MatK N.
 CC Pfam; PF01348; Intron_maturas2; 1.
 CC Pfam; PF01824; MatK N; 1.
 CC Chloroplast; mRNA processing.
 KW SEQUENCE 526 AA; 63038 MW; 4798486C56ACC011 CRC64;

Query Match 4.9%; Score 92; DB 1; Length 526;

Best Local Similarity 22.1%; Pred. No. 6.4;
 Matches 65; Conservative 35; Mismatches 118; Indels 76; Gaps 11;

QY 2 HVCERTDPKYLKIALKRLQSLSLFFONSQSLORAYSTPYVYRIILQENKEKQALAR 61
 Db 236 HVCE-----YESIFFELKRSLSHLSTSYEVLFEFIVPYGK----- 271
 QY 62 HKCISILEFFKNLL--FVHLLSLSKNQ-----RECGSTDMAVSTP-----FFNENLM 107
 Db 272 -----IHFFKVFVNFPAILGLKDPFIHYVYHGRGILLAYKTPLLMKWKYFFVNLW 326
 QY 108 YLLSSRLSKWSCPRFFLDYLEAGLLSDFLDHQAVIKFELETHFSYYPVSGFVAPH 167
 Db 327 QCYFSVWFQSQKVNINQLSKDNLEFJGYLLSURLNPLVVRSMLE-----NSFLIDN 378

QY 168 QYLSLQDRYFPIASVMTLTKDNFSLTPDLIHLGLH-----VPMLLHPSFSEFFINMR 223
 Db 379 VRIKL--DSKIPISIIIGSLAKDKFC-----NVLGHPIKATW-TDSSSDILNRFVR 428
 QY 224 LFTKVIKVOALPKQKQ-----IOTLQSNLIAIVRCFWFTVSGLE 266
 Db 429 ICRNISHYSGSGKKNLRIKYLRLCCVKTARKHKSTVRTFLKRLGSLGE 482

RESULT 46

MATK MAIZE
 ID MATK MAIZE STANDARD; PRT; 544 AA.
 AC P48190;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable intron maturase (Maturase K).
 GN MATK OR YCF14
 OS Zea mays (Maize).
 OG Chloroplast
 OC Eukaryota, Viridiplantae, Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC Clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95395841; PubMed=7666415;
 RA Maier R.M., Neckermann K., Igloi G.L., Koessel H.;
 RT "Complete sequence of the maize chloroplast genome: gene content,
 RT hotspots of divergence and fine tuning of genetic information by
 RT transcript editing.";
 RL J. Mol. Biol. 251:614-628(1995).
 CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
 CC INTRONS.
 CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS.
 CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
 CC BY MITOCHONDRIAL INTRONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X86563; CAA60266.1; -.
 CC MaizeDB; 118213; -.
 CC InterPro; IPR000442; Intron_mature2.
 CC InterPro; IPR002866; MatK N.
 CC Pfam; PF01348; Intron_maturas2; 1.
 CC Pfam; PF01824; MatK N; 1.
 CC Chloroplast; mRNA processing.
 KW SEQUENCE 544 AA; 64800 MW; 95D6C38D39EC84D0 CRC64;

Query Match 4.9%; Score 92; DB 1; Length 544;

Best Local Similarity 18.6%; Pred. No. 6.7;
 Matches 73; Conservative 64; Mismatches 140; Indels 116; Gaps 16;

QY 1 VHYCERTDPKYLKIALKRLQSLSLFFONSQSLORAYSTPYVYRIILQENKEKQALA 60
 Db 199 LEYRQDVPSSLHLRFLFYLYSNWNSFITSMKSI-----FLKKENKLPFRFL 246
 QY 61 RHKCSILEFFKNLLFVHLLS-----LSKNQREGCTDMAVSTPFFNRLWY- 108
 Db 247 YNSYVSEYEFF--LFLHKKOSSCLRLTSSGTFLERIIIFSGNHEFGVNWPGCFKTIWFF 304
 QY 109 -----RLSSRFSL-----WKSYPFRFLDYLEAGLLSDFLDHQAVIKFE 150
 Db 305 MDPLMHYVRYOGKAILASKGTLKLLKKWKSILVNFWSQVFFSWTQPORIRLNQLTNSCFD 364
 QY 151 LETHFSYYPVSGFVAPHQYLSLQDRYFPIASVMTLTKDNFSLTPDLIHLGLHVPWLL 210

DB 365 FLGYLSSVPINTLLVNRNML-----ENSLFDTRMKKFD-----ITVLATPLVGSGL----- 410
 QY 211 HSESEFPIINMGRLETK-----VIEKVAL-----PSKKRIQLTQSLI 250
 DB 411 --SKAFCCTGSGCHPSKFWTDLSMDILDRPRICRNIFHHSSSKKQLYRLK----- 464
 QY 251 AIVRCFTFTVESGLIENHGKAKYAVLSSPELGHATIDVRYLPLEDOIIRLPPT 310
 DB 465 YLLR--LSCARTLARKK-----STVRTFWRGLSVFLERP----- 498
 QY 311 STPOETLFSIRHFDLVELT-----SKLEWMLD 338
 DB 499 FTEEOVPSLM-FTKTIHFPHSGSECIMYLD 530

RESULT 47
 APB HUMAN STANDARD; PRT; 4563 AA.
 ID P04114; 000502; Q13787;
 AC 01-NOV-1986 (Rel. 03, Created)
 DT 01-NOV-1986 (Rel. 03, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Apolipoprotein B-100 precursor (Apo B-100) [Contains: Apolipoprotein B-48 (Apo B-48)].
 GN Apob.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8701385; PubMed=3763409; Pease R.J., Lusis A.J., Knott T.C., Wallis S.C., Powell L.M., Pease R.J., Lusis A.J., Blackhart B., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.; "Complete cDNA and derived protein sequence of human apolipoprotein B-100." Nucleic Acids Res. 14:7501-7503(1986).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88003974; PubMed=3652907; Ludwig E.H., Blackhart B.D., Pierotti V.R., Calaci L., Fortier C., Knott T., Scott J., Mahley R.W., Levy-Wilson B., McCarthy B.J.; "DNA sequence of the human apolipoprotein B gene." J. Biol. Chem. 261:12918-12921(1986).
 RN 13
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87008488; PubMed=3759943; Chen S.-H., Yang C.-Y., Chen P.-F., Setzer D., Tanimura M., Li W.-H., Goto A.M., Jr., Chan L.; "The complete cDNA and amino acid sequence of human apolipoprotein B-100." J. Biol. Chem. 261:12918-12921(1986).
 RN 14
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87041416; PubMed=3464946; Law S.W., Grant S.M., Higuchi K., Hospattankar A.V., Lackner K.J., Lee N., Brewer H.B., Jr.; "Human liver apolipoprotein B-100 cDNA: complete nucleic acid and derived amino acid sequence." J. Biol. Chem. 261:12918-12921(1986).
 RN 15
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87161758; PubMed=3030729; Cladaras C., Hadjopoulos-Cladaras M., Nolte R.T., Atkinson D., Zannis-Vi.; "The complete sequence and structural analysis of human apolipoprotein B-100: relationship between apob-100 and apob-48 forms." J. Biol. Chem. 261:12918-12921(1986).
 RN 16
 RP SEQUENCE OF 709-906 FROM N.A.
 RX MEDLINE=85270450; PubMed=3860836; Deeb S.S., Motulsky A.G., Albers J.J.;

RT "A partial cDNA clone for human apolipoprotein B." J. Biol. Chem. 261:12918-12921(1986).
 RN 17
 RP SEQUENCE OF 3056-3159 FROM N.A.
 RX MEDLINE=86041988; PubMed=3903660; Mehrabian M., Schumaker V.N., Fareed G.C., West R., Johnson D.F., Kirschgesner T.G., Lin H.-C., Wang X., Ma Y., Mendiaz B., Lusis A.J.; "Human apolipoprotein B: identification of cDNA clones and characterization of mRNA." J. Biol. Chem. 261:12918-12921(1986).
 RN 18
 RP SEQUENCE OF 1937-2018 AND 3811-4334 FROM N.A.
 RX MEDLINE=86093680; PubMed=3841204; Carlsson P., Olofsson S.O., Bondjers G., Darnfors C., Wiklund O., Bjursell G.; "Molecular cloning of human apolipoprotein B cDNA." J. Biol. Chem. 261:12918-12921(1986).
 RN 19
 RP SEQUENCE OF 3109-4563 FROM N.A.
 RX MEDLINE=85300528; PubMed=2994225; Knott T.J., Ball S.C., Jr., Innerarity T.L., Jacobson S.F., Uda M.S., Levy-Wilson B., Powell L.M., Pease R.J., Eddy R., Nakai H., Byers M., Priestley L.M., Robertson E., Rall L.B., Behrholz C., Shows T.B., Mahley R.W., Scott J.; "Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression, and chromosomal localization." J. Biol. Chem. 261:12918-12921(1986).
 RN 20
 RP SEQUENCE OF 1-291 FROM N.A.
 RX MEDLINE=86149325; PubMed=3513177; Procter A.A., Hardman D.A., Schilling J.W., Miller J., Appleby V., Chen G.C., Kireher S.W., McMenon G., Kane J.P.; "Isolation of a cDNA clone encoding the amino-terminal region of human apolipoprotein B." J. Biol. Chem. 261:12918-12921(1986).
 RN 21
 RP SEQUENCE OF 1-1670 FROM N.A.
 RX MEDLINE=86287319; PubMed=3461454; Procter A.A., Hardman D.A., Sato K.Y., Schilling J.W., Yamana M., Hori Y.J., Hjertild K.A., Chen G.C., Kane J.P.; "Analysis of cDNA clones encoding the entire B-26 region of human apolipoprotein B." J. Biol. Chem. 261:12918-12921(1986).
 RN 22
 RP PARTIAL SEQUENCE AND IDENTIFICATION (APO-B48).
 RX MEDLINE=88018019; PubMed=3659919; Chen S.-H., Habib G., Yang C.-H., Gu Z.-W., Lee B.R., Weng S.-H., Silbermann S.R., Gal S.-J., Deslypere J.P., Rosseneu M., Goto A.M., Jr., Li W.-H., Chan L.; "Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon." J. Biol. Chem. 261:12918-12921(1986).
 RN 23
 RP DOMAINS.
 RX MEDLINE=87039351; PubMed=3773997; Knott T.C., Pease R.J., Powell L.M., Wallis S.C., Rall S.C., Jr., Innerarity T.L., Blackhart B., Taylor W.R., Marcel Y., Milne R., Johnson D., Fuller M., Lusis A.J., McCarthy B.J., Mahley R.W., Levy-Wilson B., Scott J.; "Complete protein sequence and identification of structural domains of human apolipoprotein B." J. Biol. Chem. 261:12918-12921(1986).
 RN 24
 RP DOMAINS.
 RX MEDLINE=87161758; PubMed=3030729; Tanimura M., Li W.-H., Sparrow D.A., Dallof H., Rosseneu M., Lee P.-S., Gu Z.-W., Goto A.M., Jr., Chan L., Rosseneu M., "Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100." J. Biol. Chem. 261:12918-12921(1986).
 RN 25
 RP CALCIUM-BINDING DATA.
 RX MEDLINE=86242245; PubMed=3087360;

RA Dahti N., Lee D.M., Mok T.;
 RT "Apolipoprotein B is a calcium binding protein.";
 RL Biochem. Biophys. Res. Commun. 137:493-499(1986).
 RN [16]
 RP VARIANT SER-4338.
 RX MEDLINE=91071750; PubMed=1979313;
 RA Navajas M., Laurent A.-M., Morel J.-F., Ragab A., Cambou J.-P.,
 RA Cuny G., Cambien F., Roizes G.;
 RT "Detection by denaturing gradient gel electrophoresis of a new
 RT polymorphism in the apolipoprotein B gene.";
 RL Hum. Genet. 86:91-93(1990).
 RN [17]
 RP VARIANT FDB GLN-3527.
 RX MEDLINE=90098975; PubMed=2563166;
 RA Soria L.F., Ludwig E.H., Clarke H.R.G., Vega G.L., Grundy S.M.,
 RA McCarthy B.J.;
 RT "Association between a specific apolipoprotein B mutation and
 RT familial defective apolipoprotein B-100.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:587-591(1989).
 RN [18]
 RP VARIANT LEU-2739.
 RX MEDLINE=91016974; PubMed=2216805;
 RA Huang L.-S., Gavish D., Breslow J.L.;
 RT "Sequence polymorphism in the human apoB gene at position 8344.";
 RL Nucleic Acids Res. 18:5922-5922(1990).
 RN [19]
 RP VARIANT FDB CYS-3558.
 RX MEDLINE=95190020; PubMed=7883971;
 RA Pullinger C.R., Hennessey L.K., Chatterton J.E., Liu W., Love J.A.,
 RA Mendel C.M., Frost P.H., Malloy M.J., Schumaker V.N., Kane J.P.;
 RT "Familial ligand-defective apolipoprotein B. Identification of a new
 RT mutation that decreases LDL receptor binding affinity.";
 RL J. Clin. Invest. 95:1225-1234(1995).
 RN [20]
 RP VARIANTS L-1437; S-1914; K-2566; T-3121; A-3945; M-4128 AND T-4481.
 RX MEDLINE=97044521; PubMed=889592;
 RA Poirel O., Ricard S., Beheguy I., Souriau C., Evans A.E.,
 RA Arveiler D., Marques-Vidal P., Luc G., Roizes G., Cambien F.;
 RT "Detection of new variants in the apolipoprotein B (Apo B) gene by
 RT PCR-SSCP.";
 RL Hum. Mutat. 8:282-285(1996).
 RN [21]
 RP VARIANTS FDB GLN-3527 AND CYS-3558.
 RX MEDLINE=97403938; PubMed=9259199;
 RA Rabes J.P., Varret M., Saint-Jore B., Erlich D., Jondeau G.,
 RA Krempf M., Giraudet P., Junien C., Boileau C.;
 RT "Familial ligand-defective apolipoprotein B-100: simultaneous
 RT detection of the ARG3500-->GLN and ARG3531-->CYS mutations in a
 RT French population.";
 RL Hum. Mutat. 10:160-163(1997).
 RN [22]
 RP VARIANTS S-1914; R-1923; L-2739; D-3319; T-3427; Q-3432 AND I-3921.
 RX MEDLINE=98141125; PubMed=9490296;
 RA Leren T.P., Bakken K.S., Hoel V., Hjermann I., Berg K.;
 RT "Screening for mutations of the apolipoprotein B gene causing
 RT hypocholesterolemia.";
 RL Hum. Genet. 102:44-49(1998).
 CC -!- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
 CC CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION SIGNAL
 CC FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL PARTICLES BY
 CC THE APOB/E RECEPTOR.
 CC -!- DISEASE: DEFECTS IN APOB ARE A CAUSE OF FAMILIAL LIGAND-DEFECTIVE
 CC APOLIPOPROTEIN B-100 (FDB). IT IS A DOMINANTLY INHERITED DISORDER
 CC OF LIPOPROTEIN METABOLISM LEADING TO HYPERCHOLESTEROLEMIA AND
 CC INCREASED PRONENESS TO CORONARY ARTERY DISEASE (CAD).
 CC THE PLASMA CHOLESTEROL LEVELS ARE DRAMATICALLY ELLEVATED DUE TO
 CC IMPAIRED CLEARANCE OF LDL PARTICLES BY DEFECTIVE APOB/E RECEPTORS.
 CC -!- DISEASE: DEFECTS IN APOB ASSOCIATED WITH OTHER GENE DEFECTS
 CC (POLYGENIC) CAN BE THE CAUSE OF HYPOCHOLESTEROLEMIA.
 CC -!- MISCELLANEOUS: APO B-48 WHICH IS SYNTHESIZED ONLY BY THE
 CC INTESTINE, AND FOUND IN CHYLOMICRONS, IS A SHORTENED FORM OF APO

Query Match

4.9%; Score 92; DB 1; Length 4563;

Best Local Similarity 20.7%; Pred. No. 99;
 Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;
 QY 3 YCERLDPKYLKIALKRLQSLSPFQNSQSLQRAYSTPYSYR---IILQKENKEKQAL 59
 DB 4211 YTREELCTMFIREVGTVLQSVKSHNGSEIL-----FSYFQDLVITLPFE-----L 4257
 QY 60 ARHKCISILEFFKNLLFVHLLSLSKNORCGCTDMVSTPFFNRNLW-----YRLSS 113
 DB 4258 RKHLIDVISMYREL---LKDLSEKAEQVFKAIQSLKTKTEVL-RNLQDLLQPIFQIUED 4312
 QY 114 RESLWKSYPFRFLDYL--EAGLLSDFLDHOAVIKPFLETHSFYSYVPVSGFVAP----- 166
 DB 4313 NIKQLKEMKFTYLYNIQDEINTIFNDYIPY--VFLLKENLNLKHKFEFQNELQEA 4370
 QY 167 -----HOYLSILODRYFIASVMT-----LDKDNESLTPDLI----- 199
 DB 4371 SOELQOIHQYIMALREYFDPISIVGTVKYVELEEKIVSLKLNLLVALKDFHSEYIVSAS 4430
 QY 200 ---HDLGHVPWLHPSPSEFFINMGRFLTQVIEKQALPSKKQRIOTLOSNI 251
 DB 4431 NPTSQLSQVEQFLHRNIQIYLSILTPDGKGEKIAELSAQAQEI--IKSQAIA 4483
 RESULT 48
 ID SPEE METJA STANDARD; PRT; 293 AA.
 AC Q57761;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable spermidine synthase (EC 2.5.1.16) (Putrescine
 DE aminopropyltransferase) (SPDSY).
 GN SPEE OR M30313.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A. 2661 / ATCC 43067;
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -!- CATALYTIC ACTIVITY: S-adenosylmethionine + putrescine = 5'-
 CC methylthioadenosine + spermidine.
 CC -!- PATHWAY: FIFTH (LAST) STEP IN THE BIOSYNTHESIS OF SPERMIDINE FROM
 CC ARGinine AND METHIONINE.
 CC -!- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASES FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67486; AAB98300.1; -
 CC TIGR; M30313; -
 CC InterPro; IPR000051; SAM bind.
 CC InterPro; IPR001045; Spermine synthase.
 CC Pfam; PF01564; Spermine synth; 1.
 CC TIGRFAMs; TIGR00417; speE; 1.

DR PROSITE: PS01330, SPERMIDINE SYNTHASE; 1.
 KM Spermidine biosynthesis; transferase; Complete proteome.
 FT DOMAIN 85 126 BINDING TO DECARBOXYLATED SAM
 (POTENTIAL)
 SQ SEQUENCE 293 AA; 33899 MW; 88848C919E5452A CRC64;
 Query Match 4.8%; Score 91.5; DB 1; Length 293;
 Best Local Similarity 19.1%; Pred. No. 3.4;
 Matches 46; Conservative 36; Mismatches 72; Indels 87; Gaps 11;
 OY 157 YPVSGFVAFPHOYLSLDDRYFPIASVM---RTLDKDNFSLTPDLIHDLGHVFWLHP 212
 DB 32 YREKSGF---QEIEIIDTYDFGKALILDNTFOTTERDEF---IYHLLSHIDLPFHP 82
 OY 213 SFSEFFINMG-----RLEFKVIEKVOA-LPSKKQRIQTLSNLI 250
 DB 83 NFRNVLVIIGGGGCTVREVVHGKSVETVDFVELDEKVIETACKKMPKLSCEIDNEKXVLI 142
 OY 251 AIVRCFWFTVSSGLIENHEGRKAYGAVLISSPQELGHAFLDNVRVLPFLDQIIRLPNT 310
 DB 143 -----ITDGIKYVAETPKKDYIIVDCPPVG----- 169
 OY 311 STPOETLSIRHFDVLVLSKLEMLDQGL--ESIPLYOE-----KYL--SGFEV 359
 DB 170 --PAKGLFEKPEFYKAVFKCLN-----DDGIWQSGSEPLNLDLIONICRYLKDAGFKI 221
 OY 360 L 360
 DB 222 I 222
 RESULT 49
 RRPL_MEASE STANDARD; PRT; 2183 AA.
 ID RRPL_MEASE
 AC P35975;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
 DE (L. protein).
 GN L.
 OS Measles virus (strain Aik-C) (Subacute sclerose panencephalitis
 OS virus); sRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Morbillivirine.
 OC NCBI_TaxID=36408;
 OX 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93227570; PubMed=8470368;
 RA Mori T., Sasaki K., Hashimoto H., Makino S.,
 RT "Molecular cloning and complete nucleotide sequence of genomic RNA of
 RT the Aik-C strain of attenuated measles virus."
 RL Virus Genes 7:67-81(1993).
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
 CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).
 CC -1- SIMILARITY: WITH L. PROTEIN OF OTHER PARAMYXOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, custodian
 CC of the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC EMBL: S58435; AAB26147.1; --
 DR PIR: G48556; G48556.
 DR InterPro: IPR001016; Viral_RNA_pol_L.
 DR Pfam: PF00946; Paramyx_RNA_pol; 1.

KM Transferase; RNA-directed RNA polymerase.
 SQ SEQUENCE 2183 AA; 247754 MW; 12A686F4C6A2199 CRC64;
 Query Match 4.8%; Score 91; DB 1; Length 2183;
 Best Local Similarity 20.6%; Pred. No. 47;
 Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;
 OY 57 QALAHKCSISLEFFKNLLFVALLSLSKNREGSGSTMAVAVSPFFNRNMLRYLSSRPS 116
 DB 429 EGLTHREQVNDNMKSFAGVVF-----GCFMFLSIDSD--LTMVLYKKAALAAQR 474
 OY 117 LMKSCPRFLDYLEAFGLSDFLDHOAVIKFELTFHSPYVSGFVAFPHOYLSLQDR 176
 DB 475 EMDSVYKPEFLRYDPPKGTGS-----RLVDFVFNDSFPDYVIMVVGAV--LHDP 526
 OY 177 YFPIASVMRTLDKDNFSLTPDLIHDLGHVFWLHPFSBFFINMGRLFTKVIKVOALP 236
 DB 527 EFNL-----SYLKEKEIK-----TGRLEFAKMTYKRA-- 555
 OY 237 SKKQRIQTLSNLI-----IVRCFWFTVSSG-----LIENHEGRKAY 274
 DB 556 -----GVLENLISNGIKYFKDNGMAKDHDITKALHTLAVSGVPKDKESHRC-- 606
 OY 275 GAVLISSPQELGHAFLDNVRVLP--LELDQIIRLPNTSTPOE----- 315
 DB 607 GPVLTYSRSPVHTSTRNVRRAKGFIPGPYIRQDDTDHPENMEAVETVSAFITTDLKK 666
 OY 316 -----TLESIRHFDVLVLSKLEML--LDQGLL-----ESIPLY-- 348
 DB 667 YCLINRWYETISLFAQR-LMEIYGLPSFPQWMLHKLRTSLVLYSDPHCPDLDLAIPLXYKV 725
 OY 349 -NOE---KY-LSGFVETLCQ 362
 DB 726 PNDQIFIKYPMGIGIEGYCQ 744
 RESULT 50
 RRPL_MEASE STANDARD; PRT; 2183 AA.
 ID RRPL_MEASE
 AC P12576;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
 DE (L. protein).
 GN L.
 OS Measles virus (strain Edmonston) (Subacute sclerose panencephalitis
 OS virus).
 OC Paramyxoviridae; Paramyxovirinae; Morbillivirine.
 OC NCBI_TaxID=11235;
 OX 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88219537; PubMed=2835864;
 RA Blumberg B.M., Crowley J.C., Silverman J.I., Menonna J., Cook S.D.,
 RA Dowling P.C.;
 RT "Measles virus L protein evidences elements of ancestral RNA
 RT polymerase."
 RL Virology 164:487-497(1988).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90085790; PubMed=2596022;
 RA Cattaneo R., Schmid A., Seitelhofer P., Kaelin K., Baccaro K.,
 RA Meilen V., Padowitz J., Flanagan S., Rima B.K., Udem S.A.;
 RT "Mutated and hypermutated genes of persistent measles viruses which
 RT caused lethal human brain diseases."
 RL Virology 173:415-425(1989).
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
 CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N).

Fri Jan 10 10:56:13 2003

```
CC -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M20865; AAA46430.1; -
CC EMBL; K01711; AAA75501.1; -
CC PIR; A28919; ZLNZMV.
CC InterPro; IPR001016; Viral_RNA_pol_L.
CC Pfam; PF00946; Paramyx_RNA_pol; 1.
CC Transferase; RNA-directed RNA polymerase.
CC SEQUENCE 2183 AA; 247646 MW; 1B0B03CA2E2B6EA5 CRC64;
CC -----
Query Match 4.8%; Score 91; DB 1; Length 2183;
Best Local Similarity 20.6%; Pred. No. 47;
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;
QY 57 QALARHKCISILEEFKNLLFVHLLSLSKNQREGCSTDMVVSTPFFNRLNLYRLLSSRFS 116
Db 429 EGLTHEQCVDNWKSFAGVKP-----GCFWPLSLSD--LTWYLDKALAAALQR 474
QY 117 LMKSYCPFFLDYLEAFGLLSDFLDHOAVIKFELETHSPYPSGFGVAHQYLSLQDR 176
Db 475 EWDSVYPKEFLRYDPPKGTGS-----RRLDVFLNDSSFDYDVMYVSGAY---LHDP 526
QY 177 YFPIASVMRTLDKDNFSLTDLHDLGHVWMLHPSFSEFFINNGRIFTKVIKVCQALP 236
Db 527 EFNL-----SYSLQEKIKE-----TGRLFAMTYKMKRA-- 555
QY 237 SKQRIQTLQSNLIA-----IVRCFWFTVESG---LIENHEGRKAY 274
Db 556 -----CQVIAENLISNGIGKYGKNGMAKDEQDLTKALHTLAVSGVPKDLKESHRG--- 606
QY 275 GAVLISSPOELGHAFIDNVRLP--LELDQIIRLPNTSTPOE----- 315
Db 607 GPVLKTSRSPVHTSTRVRAAKGFIGFQVIRQODTDHPENMEAVETVSAFITTDLKK 666
QY 316 -----TLFSIRHFDLVELTSKLEWM--LDQGLL-----ESIPLY-- 348
Db 667 YCLNWRYETISLFAQR-LNEIYGLPSFQWHLKRLTSVLYVSDPHCPDLDLDAHIPLYKV 725
QY 349 -NOE---KY-LSGFVLCO 362
Db 726 PNDQIFIKYPMGGIEGYCQ 744
```

Search completed: January 9, 2003, 15:27:50
Job time : 33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2003, 11:51:03 ; Search time 65 Seconds
(without alignments)

742.103 Million cell updates/sec

Title: US-09-438-185A-1047

Perfect score: 1889

Sequence: 1 VHYCETLDPKYILKIALKL.....ESTPLNQEKYLSGFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : 1_ /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1889	100.0	362	23	AB99626
2	1250.5	66.2	259	20	AA935703
3	237	12.5	429	22	AAE11086
4	235.5	12.5	415	22	AAE11161
5	231.5	12.3	444	22	AA801192
6	227.5	12.0	444	22	AA879208
7	226.5	12.0	434	23	AA017329
8	222	11.8	498	15	AA950185
9	221	11.7	498	14	AA936741
10	220	11.6	497	15	AA950184
11	219	11.6	497	14	AA936740
12	219	11.6	497	22	AB964882
13	213.5	11.3	452	12	AA913119
14	213.5	11.3	452	18	AAW25788
15	213.5	11.3	452	20	AA95893
16	213.5	11.3	452	21	AA978593
17	206.5	10.9	532	21	AA913327
18	199.5	10.6	491	22	AA930703
19	194	10.3	452	22	AB961277
20	178.5	9.4	555	22	AB971186
21	153	8.1	200	22	AA930702
22	109.5	5.8	64	22	AAE11032
23	109.5	5.8	104	23	AB966168
24	108.5	5.7	1501	22	AB961194
25	103.5	5.5	1782	22	AB965685
26	102	5.4	463	23	AB95686
27	101	5.3	456	22	AB94081
28	101	5.3	2042	22	AB95689
29	99	5.2	524	19	AAW5848
30	98.5	5.2	1024	18	AA919604
31	98	5.2	56	22	AAE11093
32	98	5.2	689	23	AB97206
33	97	5.1	439	20	AA935368
34	95	5.0	920	21	AA931679
35	95	5.0	944	21	AA931678
36	95	5.0	970	21	AA931677
37	94	5.0	293	18	AAW55671
38	94	5.0	528	22	AA982418
39	94	5.0	1226	18	AA913825
40	93	4.9	307	23	AB955057
41	93	4.9	744	22	AB915643
42	92	4.9	333	23	AA996092
43	92	4.9	323	22	AA93268
44	92	4.9	3923	20	AA931237
45	92	4.9	4336	19	AA941262
46	92	4.9	4336	20	AA96886
47	91.5	4.8	872	18	AAW26605
48	91.5	4.8	872	20	AA931812
49	91.5	4.8	872	20	AA931813
50	91	4.8	872	20	AA931816
51	91	4.8	1081	22	AB965752
52	91	4.8	1081	22	AB965753
53	91	4.8	1081	22	AB965757
54	91	4.8	2183	19	AAW48707
55	91	4.8	2183	19	AAW48708
56	91	4.8	2183	19	AAW48709
57	91	4.8	2183	19	AAW48710
58	91	4.8	2183	19	AAW48710
59	90.5	4.8	611	23	AB909492
60	90.5	4.8	615	16	AA972726
61	90.5	4.8	615	18	AAW22226
62	90	4.8	625	23	AB939659
63	90	4.8	654	23	AB938416
64	90	4.8	892	20	AA902528
65	90	4.8	2183	14	AA939592
66	90	4.8	4590	22	AA933184
67	90	4.8	4591	22	AB932977
68	89	4.7	392	19	AAW33819
69	89	4.7	392	21	AA926803
70	89	4.7	392	22	AB92269
71	89	4.7	392	23	AAE13669
72	89	4.7	2055	22	AB970206
73	88.5	4.7	320	22	AA971806
74	88.5	4.7	801	22	AB958990
75	88	4.7	332	22	AB911297
76	88	4.7	410	22	AB903127
77	88	4.7	438	18	AAW40373
78	88	4.7	438	18	AAW01572
79	88	4.7	438	21	AA94036
80	88	4.7	641	22	AB958034
81	88	4.7	659	21	AA94047
82	88	4.7	720	21	AA94049
83	88	4.7	749	21	AA94048

Human tyrosine hyd
Drosophila melanog
Phenylalanine hydr
Human phenylalanin
Human phenylalanin
Human phenylalanin
Caenorhabditis ele
Novel human secret
Drosophila melanog
Drosophila melanog
Novel human secret
phAB fusion pepti
Human ORFX protein
Drosophila melanog
Drosophila melanog
Fungal ZAC protein
Human protein sequ
Drosophila melanog
Amino acid sequenc
Mycoplasma genital
phAB fusion pepti
Novel human protei
C. pneumoniae prot
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
H. pylori ORF hp5p
S. epidermidis ope
Yeast transcriptio
Lactococcus lactis
Novel human diagn
RhCMV unique short
Novel human secret
Human Apo B protei
Apollipoprotein B-1
Amino acid sequenc
Thermoanaerobacter
Thermoanaerobacter
T. thermohydrosulf
Drosophila melanog
Drosophila melanog
Drosophila melanog
Measles virus Rube
Measles virus Mora
Measles virus Zag
Measles virus Aik
Measles virus Edmo
AmEPV subunit of ye
ORC3 subunit of ye
S. cerevisiae orig
Staphylococcus epi
Staphylococcus epi
Amino acid sequenc
L protein of atten
Novel human secret
Novel human diagn
Pseudomonas Oris
Pseudomonas alcali
Pseudomonas alcali
Drosophila melanog
Human olfactory re
Drosophila melanog
Drosophila melanog
Novel human diagn
Novel human diagn
Human breast cance
Protein encoded by
Amino acid sequenc
Drosophila melanog
Amino acid sequenc
Splice variant of

Db 61 RHKISILEFFKNLLFVHLLSLSKNQREGSCDTMAVSTPFFNRLWYRLSSRFLWKS 120
 QY 121 YCPFFFDYLEAFGLLSDFLDHQAIVKFFELTHPSYYPVSGFVAPHOYLSLLQDRYFPI 180
 Db 121 YCPFFFDYLEAFGLLSDFLDHQAIVKFFELTHPSYYPVSGFVAPHOYLSLLQDRYFPI 180
 QY 181 ASVVRTLDKDNFSLTPDLIHDLGHVPLHPSFSEFFINNGRLFTKVIKVOALPSKKQ 240
 Db 181 ASVVRTLDKDNFSLTPDLIHDLGHVPLHPSFSEFFINNGRLFTKVIKVOALPSKKQ 240
 QY 241 RIOTLQSNLTAIVRCFWFTVESGLIENHGRKAYCAVLISSPOELGHAFIDNVRLPLEL 300
 Db 241 RIOTLQSNLTAIVRCFWFTVESGLIENHGRKAYCAVLISSPOELGHAFIDNVRLPLEL 300
 QY 301 DOIIRLPNTSTPQETLFSIRHDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL 360
 Db 301 DOIIRLPNTSTPQETLFSIRHDELVELTSKLEWMLDQGLLESIPLYNQEKYLSGFEVL 360
 QY 361 CQ 362
 Db 361 CQ 362

RESULT 2
 AAY35703
 ID AAY35703 standard; Protein; 259 AA.

AC AAY35703;
 DT 13-SEP-1999 (first entry)
 XX Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.
 XX WO9927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (G8ST) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 1410-1411; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX Sequence 259 AA;

Query Match 66.2%; Score 1250.5; DB 20; Length 259;
 Best Local Similarity 91.8%; Pred. No. 1.3e-117;
 Matches 245; Conservative 3; Mismatches 8; Indels 11; Gaps 1;
 QY 96 VVSTPFFNRLWYRLSSRFLWKSVCPRFPLDYLLEAFGLLSDFLDHQAIVKFFELTHF 155
 Db 4 MVSTPFLTVMSEKLLS-----KIFLDYLEAFGLLSDFLDHQAIVKFFELTHF 52
 QY 136 SYYPVSGFVAPHOYLSLLQDRYFPTASVMRTLDKDNFSLTPDLIHDLGHVPLHPSFS 215
 Db 53 SYYPVSGFVAPHOYLSLLQDRYFPTASVMRTLDKDNFSLTPDLIHDLGHVPLHPSFS 112
 QY 216 EFFINMGRLFTKVIKVOALPSKKQRIOTLQSNLTAIVRCFWFTVESGLIENHGRKAYG 275
 Db 113 EFFINMGRLFTKVIKVOALPSKKQRIOTLQSNLTAIVRCFWFTVESGLIENHGRKAYG 172
 QY 276 AVLSSPOELGHAFIDNVRLPLELDOIIRLPNTSTPQETLFSIRHDELVELTSKLEW 335
 Db 173 AVLSSPOELGHAFIDNVRLPLELDOIIRLPNTSTPQETLFSIRHDELVELTSKLEW 232
 QY 336 MLDQGLLESIPLYNQEKYLSGFEVLQ 362
 Db 233 MLDQGLLESIPLYNQEKYLSGFEVLQ 259

RESULT 3
 AAE11086
 ID AAE11086 standard; Protein; 429 AA.
 AC AAE11086;
 DT 18-DEC-2001 (first entry)
 XX Protein encoded by pFUSAB vector DNA insert sequence.

XX Phenylalanine hydroxylase; PAH; phhA; 4a-carbinolamine dehydratase; phhB;
 KW aromatic aminotransferase; phhC; milk protein; animal protein; casein;
 KW proteinaceous food product; globulin; whey protein; phenylketonuria;
 KW PKU; inherited metabolic disorder; impaired brain function; nootropic;
 KW cell therapy.

XX Chimeric - Escherichia coli.
 OS Chimeric - Unidentified.

Key	Location/Qualifiers
FT Region	1..26
FT Region	/label= LacZ_peptide
FT Region	27..28
FT Region	/label= Linker_peptide
FT Region	29..289
FT Region	/label= phhA protein
FT Region	290..311
FT Region	/label= Junction_peptide
FT Region	312..429
FT Region	/label= phhB protein

XX WO200168822-A2.

XX 20-SEP-2001.

XX 14-MAR-2001; 2001WO-DK00172.

XX 14-MAR-2000; 2000US-0525116.

XX (NILA-) NILAB APS.

XX Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L;

PI Arnau J, Jensen SH, Gjetting T, Nielsen E;

CC WPI; 2001-590055/66.

XX N-PSDB; RAD18533.

XX Novel recombinant cells comprising a nucleic acid encoding a gene

product having phenylalanine hydroxylase activity, that is derived from a prokaryotic organism, is useful for treating phenylketonuria in mammals -

Example 2; Fig 5; 91pp; English.

The patent discloses novel cells comprising a nucleic acid encoding a gene product having phenylalanine hydroxylase (PAH) activity such as phenylalanine hydroxylase (paha), 4a-carbinolamine dehydratase (phhb) and aromatic aminotransferase (phbc), which are derived from a prokaryotic organism. The patent also relates to fusion proteins comprising a protein enhancing and/or stabilising the PAH activity in addition to PAH activity. The cells are useful for producing PAH. The sequences of the invention are also useful as a milk protein derived from casein, globulin or a whey protein) having reduced content of phenylalanine. The method involves contacting the food product starting material with the cells or fusion proteins such that at least part of the phenylalanine content of the starting material is converted into compounds that do not cause phenylketonuria (PKU) by the enzymatically active product. PAH enzyme is useful for manufacturing a medicament for treating PKU, which is an inherited metabolic disorder resulting in an accumulation in the body of L-phenylalanine and metabolites that can cause impaired brain function. The present sequence is a protein encoded by pFUSAB vector DNA insert sequence. This sequence comprises a lacZ peptide from Escherichia coli, a linker peptide and paha and phhb proteins.

Sequence 429 AA;

Query Match 12.5%; Score 237; DB 22; Length 429;

Best Local Similarity 23.9%; Pred. No. 4,4e-15;

Matches 66; Conservative 56; Mismatches 114; Indels 40; Gaps 7;

66 SLEPFKMLFVHLSTSKQRCSTDMANVSTPPNNLWYLLSRFLMSYCFR 125
 20 SVLEFGFKLTQTVARQPD-----DNGFIHYPETHEQVNTLITQLKVIGRACOE 72
 126 FLVDLEAFGLSDFLDHQAVIKFFELE-----THSSYYPVSGFAFHQVYSLIQDVPF 119
 73 YLDELEQUG-----LPHERIPQDELRNVLAQTGWRVAVPALIPOTFFELLASQPF 127
 180 IASVWRTLDKDNFSTPDLHDHGLGVHPLHSPSEPPINNGRLFTTVIKVQALSKK 239
 128 VAIIRIPESLDITQEPDITHEIFGCPILNFWFAEFTHTYKGLKA-----SKK 179
 240 QRIOTLOSNIATVRCFWFVESGLIENHEGRKANGAVLISSPOELGHAFD--NVRVL 296
 180 ER-----VFARLYWMTIEBGLVETQGRKRYGGGLSSPKETVYLSDEPLHQAEN 231
 297 PLELDQIIRLPENTSTPQETLFSIRHPELVEVTSK 332
 232 PLE--AMRTFVRIDLIQPLFVLPDLKRLFOLAQE 264

RESULT 4

AAE11161 standard; Protein; 415 AA.

AAE11161;

18-DEC-2001 (first entry)

phhAB fusion protein fragment.

Phenylalanine hydroxylase; PAH; paha; 4a-carbinolamine dehydratase; phhb;

aromatic aminotransferase; phbc; milk protein; animal protein; casein;

proteineous food product; globulin; whey protein; phenylketonuria;

PKU; inherited metabolic disorder; impaired brain function; noctropic;

celi therapy; phhAB fusion protein.

Identified.

Key Location/Qualifiers
 Region 21..46 /note= "phhAB fusion peptide #1"
 Region 55..90 /note= "phhAB fusion peptide #2"
 Region 94..157 /note= "phhAB fusion peptide #3"
 Region 168..223 /note= "phhAB fusion peptide #4"
 Region 228..272 /note= "phhAB fusion peptide #5"
 Region 295..336 /note= "phhAB fusion peptide #6"
 Region 353..389 /note= "phhAB fusion peptide #7"
 Region 396..415 /note= "phhAB fusion peptide #8"

W0200168822-A2.

20-SEP-2001.

14-MAR-2001; 2001WO-DK00172.

14-MAR-2000; 2000US-0525116.

(NIIA-) NIIAB AFS.

Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L;

Arnaud J, Jensen SH, Gjetting T, Nielsen E;

WPI; 2001-590055/66.

Novel recombinant cells comprising a nucleic acid encoding a gene product having phenylalanine hydroxylase activity, that is derived from a prokaryotic organism, is useful for treating phenylketonuria in mammals -

Example 8; Fig 18; 91pp; English.

The patent discloses novel cells comprising a nucleic acid encoding a gene product having phenylalanine hydroxylase (PAH) activity such as phenylalanine hydroxylase (paha), 4a-carbinolamine dehydratase (phhb) and aromatic aminotransferase (phbc), which are derived from a prokaryotic organism. The patent also relates to fusion proteins comprising a protein enhancing and/or stabilising the PAH activity in addition to PAH activity. The cells are useful for producing PAH. The sequences of the invention are also useful as a milk protein derived from casein, globulin or a whey protein) having reduced content of phenylalanine. The method involves contacting the food product starting material with the cells or fusion proteins such that at least part of the phenylalanine content of the starting material is converted into compounds that do not cause phenylketonuria (PKU) by the enzymatically active product. PAH enzyme is useful for manufacturing a medicament for treating PKU, which is an inherited metabolic disorder resulting in an accumulation in the body of L-phenylalanine and metabolites that can cause impaired brain function. The present sequence is phhAB fusion protein fragment.

Sequence 415 AA;

Query Match 12.5%; Score 235.5; DB 22; Length 415;

Best Local Similarity 24.5%; Pred. No. 5,9e-15;

Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;

93 DMAVSTPPNNLWYLLSRFLMSYCFRFLVDLEAFGLSDFLDHQAVIKFFELE 152

26 DNGFIHYPETHEQVNTLITQLKVIGRACOEFLDITQEPDITHEIFGCPILNFWFAEFTHTYKGLKA-----LPHERIPQDELRNV 80

153 -----THSSYYPVSGFAFHQVYSLIQDVPF 206

81 RVLDQIIRLPENTSTPQETLFSIRHPELVEVTSK 140


```

QY 207 PILLHPSSEFFINMGRLETKVIEKQALPSKQRIQTLQSNLIAIVRCFWFTVSGLE 266
DB 141 PLLTNWFAEFTHYKGLKA-----SKER-----VFLARYWNTIEFGLVE 184
QY 267 NHEGRKAYGNVLSSPOELGHAFID---NVRVLELDQIIRLPFTNTSTPQETLSIRHF 323
DB 185 TDQKRIYGGIUSPKEIVISLSDEPLHQFNLE---AWRTPIRIDILQPLVFLPDL 241
QY 324 DELVELTSK 332.
DB 242 KRUFOLAE 250

RESULT 5
ID AAM80192
AC AAM80192 standard; Protein; 444 AA.
XX
XX AAM80192;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3838.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX 20-JUN-2000; 2000US-0598075.
XX
XX 19-JUL-2000; 2000US-0620325.
XX
XX 01-SEP-2000; 2000US-0654936.
XX
XX 15-SEP-2000; 2000US-0663561.
XX
XX 20-OCT-2000; 2000US-0693325.
XX
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAKS3325.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 446-447; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAAM80302) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and the polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing

```

```

CC were missing at the time of publication.
SQ Sequence 444 AA;
Query Match 12.3%; Score 231.5; DB 22; Length 444;
Best Local Similarity 27.8%; Pred. No. 1.6e-14;
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;
QY 107 WYRLSSRPSLWKSYPFFLDYLEAFGLLSDFLDH-----QAVIKFPELETHFSY 158
DB 174 WGTVFQELNKLYPTHACR---EYLNKLPLLSKYCYGREDNIPQLEDVSNFLKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLQDQRYFPIASVMRTLDKDNFSLTPDLIHLGLHVPMLHPSPSEPF 218
DB 231 PVAGYLSPRDFLSGLAFRVFCHTQYVRHSSDPFYTPBEDTCHELLGHVPLLAEPSPAFQS 290
QY 219 INMGRLETKVIEKQALPSKQRIQTLQSNLIAIVRCFWFTVSGLEHGRKAYGAVL 278
DB 291 QEIG-----LASLGASEAVQKLAT-----CYFFTVEGLCKDQGLRVFAGL 334
QY 279 ISSPQELGHAFIDNVRVLELDQIIRLPFTNTSTPQETLSIRHDELVE 328
DB 335 LSSISELKHLSGHARVKEPDKITCKOECLITTFQDVYFVSVSEFEDAKE 384

RESULT 6
ID AAM79208
AC AAM79208 standard; Protein; 444 AA.
XX
XX AAM79208;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 1870.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX 20-JUN-2000; 2000US-0598075.
XX
XX 19-JUL-2000; 2000US-0620325.
XX
XX 01-SEP-2000; 2000US-0654936.
XX
XX 15-SEP-2000; 2000US-0663561.
XX
XX 20-OCT-2000; 2000US-0693325.
XX
XX 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAKS2341.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 4255-4256; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAAM80302) that exhibit activity relating to

```

CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulatory
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAK80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX Sequence 444 AA;

Query Match 12.0%; Score 227.5; DB 22; Length 444;
 Best Local Similarity 27.4%; Pred. No. 4,2e-14;
 Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

OY 107 MYRLSSRFSIMKSCYCRFLDYLEAFGLSDPFDH-----QAVIKFPELETHFSY 158
 DB 174 WGTFOELNKLYPTHACK---EYLKMLPLSKICGIRENIPQLEDVSNFLKERTGFSIR 230
 OY 159 PVSGFVAAPHOYLILQDRFPFIASVMRTLDKDNFSLTPDLIHLLGHVFWLLHPSSEFF 218
 DB 221 PVAGYLSRDFLSGLAFRFVHCTQYVRHSSDPYTPPEPTCHELGHVPLLAEPSPAFQS 290
 OY 219 INMGRLEFTKYLEKVALPSKKORIQTLOSNIILAVRCFMFTVESGLIENHGRKAYGAVL 278
 DB 291 OEIG-----LASLGASEAVQKLTAT-----CYFTVEFGLCQDQGLRVFGAGL 324
 OY 279 ISSFOELGHAFIDNVRVLPLELDQIRLPFNTSTPOETLFSIRHDELVE 328
 DB 335 LSSISELKHVLSGHAKVFPDPKITCKQECILTFQDVFVSESPEDAKE 384

RESULT 7

ID AAO17329 standard; Protein; 434 AA.

XX AAO17329;

DT 08-JUL-2002 (first entry)

DE Neurone specific tryptophane hydroxylase.

XX Neuronal tryptophane hydroxylase; TPH; neuronal disease;

KW primary haemostasis deficiency; allergy; transplantation; serotonin;

KW antiallergic; gynaecological.

XX Unidentified.

PN WO200217891-A2.

PD 07-MAR-2002.

PF 27-AUG-2001; 2001WO-DE03178.

PR 31-AUG-2000; 2000DE-1043124.

PA (DELB-) DELBRUECK CENT MOLEKULARE MEDIZIN MAX.

PI Walther D. Bader M;

DR WPI; 2002-281015/32.

PT Regulating levels of serotonin, useful for diagnosing neuronal disease

PT and treating primary hemostatic deficiency, comprises controlling

PT activity of tryptophan hydroxylase -

PS Claim 21; Fig 9; 21pp; German.

CC The present invention relates to a method of influencing the level of
 CC serotonin, which involves the specific regulation of TPH (tryptophan
 CC hydroxylase) and/or neurone-specific TPH (nrTPH) activity. The method is
 CC used for the diagnosis of neuronal diseases and for treating deficient
 CC primary haemostasis. Also, reducing production of serotonin is used to
 CC treat arteriosclerosis and thrombosis, particularly where associated with
 CC diabetes mellitus, and to treat excessive (or normal but unwanted) immune
 CC system responses, e.g. allergy, (auto)immune diseases, risks associated
 CC with pregnancy, particularly pre-eclampsia, and diseases, risks associated
 CC present sequence is a neurone-specific TPH protein.

XX Sequence 434 AA;

Query Match 12.0%; Score 226.5; DB 23; Length 434;
 Best Local Similarity 27.4%; Pred. No. 5.1e-14;
 Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

OY 107 MYRLSSRFSIMKSCYCRFLDYLEAFGLSDPFDH-----QAVIKFPELETHFSY 158
 DB 164 WGTFOELNKLYPTHACK---EYLKMLPLSKICGIRENIPQLEDVSNFLKERTGFSIR 220
 OY 159 PVSGFVAAPHOYLILQDRFPFIASVMRTLDKDNFSLTPDLIHLLGHVFWLLHPSSEFF 218
 DB 221 PVAGYLSRDFLSGLAFRFVHCTQYVRHSSDPYTPPEPTCHELGHVPLLAEPSPAFQS 280
 OY 219 INMGRLEFTKYLEKVALPSKKORIQTLOSNIILAVRCFMFTVESGLIENHGRKAYGAVL 278
 DB 281 OEIG-----LASLGASEAVQKLTAT-----CYFTVEFGLCQDQGLRVFGAGL 324
 OY 279 ISSFOELGHAFIDNVRVLPLELDQIRLPFNTSTPOETLFSIRHDELVE 328
 DB 325 LSSISELKHVLSGHAKVFPDPKITCKQECILTFQDVFVSESPEDAKE 374

RESULT 8

ID AAR50185 standard; Protein; 498 AA.

XX AAR50185;

DT 04-NOV-1994 (first entry)

DE Rat tyrosine hydroxylase.

XX Human, rat; tyrosine hydroxylase; TH; substitution; N-terminal;

KW enhanced enzymatic activity; wild-type; transformation; dopamine;

KW Parkinson's disease; Alzheimer's disease; brain; encapsulation;

XX selectively permeable; polymer capsule; antibody.

OS Rattus rattus.

PN US5300436-A.

PD 05-APR-1994.

PF 13-MAR-1991; 91US-0669446.

PR 13-MAR-1991; 91US-0669446.

PR 26-JAN-1993; 93US-0009075.

PA (UYNV) UNIV NEW YORK STATE.

PI Filler D, Friedhoff AJ, Goldstein M, Wu J;

DR WPI; 1994-125849/15.

PT Modified DNA encoding variant tyrosine hydroxylase with an

PT N-terminal amino acid substn. and cells contg. it - can be used

PT enzyme, e.g. Parkinson's disease or Alzheimer's disease

PS Disclosure; Fig 5; 27pp; English.

CC This sequence represents rat tyrosine hydroxylase (TH) and may be
 CC used in the production of the variant TH molecules of the invention.
 CC These variants contain amino acid substitutions in the N-terminal
 CC portion of the molecule, and in the rat molecule these substitutions
 CC are pref. at positions Ser8, Ser19, Ser31, Arg38, Ser40, Glu43 or
 CC Arg46. Variants containing one or more of these amino acid
 CC substitutions, have substantially enhanced enzymatic activity
 CC compared to the wild-type enzyme. The most pref. substitution being
 CC Ser40 for Tyr or Leu. DNA encoding the TH variants, and cells
 CC transformed with this DNA may be used for treating diseases associated
 CC with defective function of TH, or dopamine, eg. Parkinsons disease
 CC and Alzheimers disease, and effective disorders. The cells can then
 CC be implanted into the brain or encapsulated in a selectively permeable
 CC polymer capsule which allows release of the cells products but protects
 CC them from attack by the hosts antibodies or cells.

XX Sequence 498 AA;

Query Match 11.8%; Score 222; DB 15; Length 498;
 Best Local Similarity 26.7%; Pred. No. 1.8e-13;
 Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRFSLMKS-----YCRFPLDYLEAFGLLSDFLDH-----QAVIKFELETHF 155
 DB 227 ABEATATWKEVVTIKGLYATHACREHLEGFQLLERYCGYREDSPQLEDVSRFLKERTGF 286
 QY 156 SYYPVSGFVAPHQVLSLLODRYPPIASVNRDLTKDNFSLTPDLIHDLGHVPMLLHPSPS 215
 DB 287 QLRFPVAGLLSAKDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMADRTFA 346
 QY 216 EEFINMGRLFKTVIEKQVLPKQKRIQTLOSNIATVRCFWFTVESGLIENHEGRKAYG 275
 DB 347 QFSQDIG-----LASIGASDEIEKLSI-----VVFVTFEFLCKQNGELKAYG 390
 QY 276 AVLISSQPELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHF-----DELVELTS 331
 DB 391 AGLSSYGEHLSLSEBEVRAFPDPTAAVQYQDOTYQPVYFVSEFNDKDKLRNYS 450
 QY 332 KLE 334
 DB 451 RIQ 453

RESULT 9
 AAR36741
 ID AAR36741 standard; protein; 498 AA.

AC AAR36741;
 DT 20-SEP-1993 (first entry)
 DE Rat tyrosine hydroxylase.
 KW Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;
 KW neurological disorder; antibody; variant.
 OS Rattus rattus.

Key Location/Qualifiers
 FT Modified-site 8 /note= "Phosphorylation site"
 FT Modified-site 19 /note= "Phosphorylation site"
 FT Modified-site 40 /note= "Phosphorylation site"
 FT Modified-site 153 /note= "Phosphorylation site"
 FT Modified-site /note= "Phosphorylation site"

XX US5212082-A.
 XX 18-MAY-1993.
 XX 13-MAR-1991; 91US-0669446.

XX 13-MAR-1991; 91US-0669446.
 XX (UUNY) UNIV NEW YORK STATE.
 XX Filler D, Friedhoff AJ, Goldstein M, Wu J;
 XX WPI; 1993-175456/21.
 XX Genetically modified tyrosine hydroxylase having increased
 XX activity - used for treating neurological disorders e.g.
 XX Parkinson's and Alzheimer's diseases and affective disorders
 XX Disclosure; Fig 5; 20pp; English.
 XX The sequences given in AAR36740-41 represent the human and rat
 XX tyrosine hydroxylase protein respectively. These sequences may be
 XX used in the construction of a variant tyrosine hydroxylase which
 XX contains at least one amino acid substitution in the N-terminal 55
 XX amino acids. The substitution corresponds to an amino acid selected
 XX from Ser8, Ser31, Arg37, Arg38, Gln39, Ser40, Leu42, Ile42, Glu43,
 XX Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant
 XX proteins is at least 3-fold greater than that for the wild type
 XX protein. Cells transfected with the DNA encoding these proteins may
 XX be used for treating neurological disorders associated with a
 XX deficiency in tyrosine hydroxylase or dopamine. These proteins may
 XX be used to generate antibodies specific for the variant tyrosine
 XX hydroxylases to monitor the enzyme during a treatment regimen.

XX Sequence 498 AA;

Query Match 11.7%; Score 221; DB 14; Length 498;
 Best Local Similarity 27.4%; Pred. No. 2.2e-13;
 Matches 65; Conservative 37; Mismatches 99; Indels 36; Gaps 5;

QY 118 WKS-----YCRFPLDYLEAFGLLSDFLDH-----QAVIKFELETHFSYPYS 161
 DB 233 WKEVVTIKGLYATHACREHLEGFQLLERYCGYREDSPQLEDVSRFLKERTGFQLRPVA 292
 QY 162 GFVAPHQVLSLLODRYPPIASVNRDLTKDNFSLTPDLIHDLGHVPMLLHPSPSEFFINM 221
 DB 293 GLLSAKDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMADRTFAQFSQDI 352
 QY 222 GRLEFTKVIKQVLPKQKRIQTLOSNIATVRCFWFTVESGLIENHEGRKAYGAVLISS 281
 DB 353 G-----LASIGASDEIEKLSI-----VVFVTFEFLCKQNGELKAYGAGILSS 396
 QY 282 PQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHF-----DELVELTSKLE 334
 DB 397 YGELLHLSLSEBEVRAFPDPTAAVQYQDOTYQPVYFVSEFNDKDKLRNYSRIQ 453

RESULT 10
 AAR50184
 ID AAR50184 standard; Protein; 497 AA.

AC AAR50184;
 DT 04-NOV-1994 (first entry)
 DE Human tyrosine hydroxylase.
 KW Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;
 KW enhanced enzymatic activity; wild-type; transformation; dopamine;
 KW Parkinsons disease; Alzheimers disease; brain; encapsulation;
 KW selectively permeable; polymer capsule; antibody.
 OS Homo sapiens.
 XX US5300436-A.
 XX 05-APR-1994.

PF 13-MAR-1991; 91US-0669446.
 XX 13-MAR-1991; 91US-0669446.
 PR 26-JAN-1993; 93US-0009075.
 XX (UNYV) UNIV NEW YORK STATE.
 PA
 XX Filer D, Friedhoff AJ, Goldstein M, Wu J;
 PI WPI; 1994-125849/15.
 DR N-PSDB; AAQ44699.
 XX
 XX Modified DNA encoding variant tyrosine hydroxylase with an
 PT N-terminal amino acid subseq. and cells contg. it - can be used
 PT to treat diseases associated with defective function of the
 PT enzyme, e.g. Parkinson's disease or Alzheimer's disease
 XX
 XX
 PS Disclosure; Column 21-24; 27pp; English.
 XX
 XX This sequence represents human tyrosine hydroxylase (TH) and may be
 CC used in the production of the variant TH molecules of the invention.
 CC These variants contain amino acid substitutions in the N-terminal
 CC portion of the molecule, and in the human molecule these
 CC substitutions are pref. at positions Ser19, Ser31, Arg37, Arg38,
 CC Ser40, Leu41, Glu43, Asp44, Ala45, Arg46 or Lys47. Variants
 CC containing one or more of these amino acid substitutions, have
 CC substantially enhanced enzymatic activity compared to the wild-type
 CC enzyme. The most pref. substitution being Ser40 for Tyr or Leu.
 CC DNA encoding the TH variants, and cells transformed with this DNA
 CC may be used for treating diseases associated with defective function
 CC of TH, or dopamine, eg. Parkinsons disease and Alzheimers disease,
 CC and effective disorders. The cells can then be implanted into the
 CC brain or encapsulated in a selectively permeable polymer capsule
 CC which allows release of the cells products but protects them from
 CC attack by the hosts antibodies or cells.
 XX
 XX
 SQ Sequence 497 AA;
 Query Match 11.6%; Score 220; DB 15; Length 497;
 Best Local Similarity 26.7%; Pred. No. 2, 8e-13;
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;
 QY 112 SSRFSLMKS-----YCPFFLDYEAFGLLSPFLDH-----QAVIKFFLETHP 155
 DB 226 AEELATWKEVYTTLKGIVATHACGHELEAFALERSGVRDNIPOLEDVSFRLEKRTGF 285
 QY 156 SYYPVSGFVAPHQVLSLLQDRYFPPIASVMTLKDNPSTLPDLIHDLGHVPLLPSPFS 215
 DB 286 QLRPVAGLLSARDPLASLAFVFOCTQYIRHASSPMHSPPDCCHLLGHVPLMLADRTFA 345
 QY 216 EEFINMGRLLFTVKYIEKVALPSKKORIQTOSNLIAIVRCFMTVESGLIENHEGRKAYG 275
 DB 346 QFSQDIG-----IASIGASDEIEKLTSLT-----WTFVEFGCKONGEVKAYG 389
 QY 276 AVLISSPOELGHAFFIDNVRLPLELDQIRLPFNTSTPOETLFSIRNF-----DELVELTS 331
 DB 390 AGLISYGEILHCLSEEPETIRAFDEPAAVOPQDOTQSVYVSESFSFSDAKDKLSRYAS 449
 QY 332 KLE 334
 DB 450 RIQ 452
 RESULT 11
 AAR36740
 ID AAR36740 standard; protein; 497 AA.
 XX AAR36740;
 AC AAR36740;
 XX 20-SEP-1993 (first entry)
 DT Human tyrosine hydroxylase.
 XX
 DE

KW Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;
 XX neurological disorder; antibody; variant.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Modified-site 19
 FT Modified-site /note= "Phosphorylation site"
 FT Modified-site 40
 FT Modified-site /note= "Phosphorylation site"
 FT Modified-site 152
 FT Modified-site /note= "Phosphorylation site"
 FT Modified-site 222
 FT /note= "Phosphorylation site"
 PN US5212082-A.
 XX
 PD 18-MAY-1993.
 XX
 XX 13-MAR-1991; 91US-0669446.
 PF 13-MAR-1991; 91US-0669446.
 PR 13-MAR-1991; 91US-0669446.
 XX (UNYV) UNIV NEW YORK STATE.
 PA
 XX Filer D, Friedhoff AJ, Goldstein M, Wu J;
 PI WPI; 1993-175456/21.
 DR N-PSDB; AAQ41886.
 XX
 XX Genetically modified tyrosine hydroxylase having increased
 PT activity - used for treating neurological disorders e.g.
 PT Parkinson's and Alzheimer's diseases and affective disorders
 XX
 XX
 PS Disclosure; Fig 6; 20pp; English.
 XX
 XX The sequences given in AAR36740-41 represent the human and rat
 CC tyrosine hydroxylase protein respectively. These sequences may be
 CC used in the construction of a variant tyrosine hydroxylase which
 CC contains at least one amino acid substitution in the N-terminal 55
 CC amino acids. The substitution corresponds to an amino acid selected
 CC from Ser9, Ser31, Arg37, Arg38, Ser40, Leu42, Leu43, Glu43,
 CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant
 CC proteins is at least 3-fold greater than that for the wild type
 CC protein. Cells transfected with the DNA encoding these proteins may
 CC be used for treating neurological disorders associated with a
 CC deficiency in tyrosine hydroxylase or dopamine. These proteins may
 CC be used to generate antibodies specific for the variant tyrosine
 CC hydroxylases to monitor the enzyme during a treatment regimen.
 XX
 XX
 SQ Sequence 497 AA;
 Query Match 11.6%; Score 219; DB 14; Length 497;
 Best Local Similarity 26.7%; Pred. No. 3, 5e-13;
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;
 QY 112 SSRFSLMKS-----YCPFFLDYEAFGLLSPFLDH-----QAVIKFFLETHP 155
 DB 226 AEELATWKEVYTTLKGIVATHACGHELEAFALERSGVRDNIPOLEDVSFRLEKRTGF 285
 QY 156 SYYPVSGFVAPHQVLSLLQDRYFPPIASVMTLKDNPSTLPDLIHDLGHVPLLPSPFS 215
 DB 286 QLRPVAGLLSARDPLASLAFVFOCTQYIRHASSPMHSPPDCCHLLGHVPLMLADRTFA 345
 QY 216 EEFINMGRLLFTVKYIEKVALPSKKORIQTOSNLIAIVRCFMTVESGLIENHEGRKAYG 275
 DB 346 QFSQDIG-----IASIGASDEIEKLTSLT-----WTFVEFGCKONGEVKAYG 389
 QY 276 AVLISSPOELGHAFFIDNVRLPLELDQIRLPFNTSTPOETLFSIRNF-----DELVELTS 331
 DB 390 AGLISYGEILHCLSEEPETIRAFDEPAAVOPQDOTQSVYVSESFSFSDAKDKLSRYAS 449
 QY 332 KLE 334

```

Db 450 RIQ 452
:::
RESULT 12
ABB64882
ID ABB64882 standard; Protein; 579 AA.
AC ABB64882;
XX
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 21438.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL08985.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 21438; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABB01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 579 AA;
SQ
Query Match 11.6%; Score 219; DB 22; Length 579;
Best Local Similarity 28.6%; Pred. No. 4, 4e-13;
Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;
OY 128 DYLEAFGLSD---FLDH-----QAVIKFELETHFSYVPSGVPAHQVLSLLODRYFP 179
Db 329 EYRAAFQKLQDEQIFVETRLPOLQNSDFLRKNTGFSLRPAAGLTTARDFLASLRIFQ 388
OY 180 TASVWRLDKNSLTPDLIHLHLLGHVPMLLHPSPSEFFINMGRFLTFTYKVKQALPSKK 239
Db 389 STQVVRHNSYHTPEPDSIHELGHMPLLDLAPSPAQSQBIG-----LASLGASD 439
OY 240 QRIOFLOSNIATVRCFFWTFSEGLIENHEGRKAYCAVLISSPOELGHAFID----- 291
Db 440 BEIEKLST-----VYWFTEFGLCKEKGQIKAYCAGLLSSYGELHLAISDKCEHRAFE 492
OY 292 --NVRVLPLE-----LQILRLPNTSTPQETLFSIRHFD--ELVELT 330
:::

```

```

Db 493 PASTAVQYQDQEQYQIYYVAESFEDAKKFRRWVSTWSPRFEVRFN-PHTERVEVLDSV 551
OY 331 SKLEWMLDQGLLESIPLYN 349
:::
Db 552 DKLETLVHMOMTEILHLTN 570
:::
RESULT 13
AAR13119
ID AAR13119 standard; Protein; 452 AA.
XX
XX AAR13119;
AC
XX
XX 08-OCT-1991 (first entry)
DT
XX
XX Phenylalanine hydroxylase.
DE
XX
XX Hybrid; fusion; membrane translocation; binding region; HIV;
KW infection; toxin; steroid; hormone; monoclonal antibody; antigen;
KW diphtheria; exotoxin; phenylketonuria; cholera; interleukin; IL-2;
KW protease; epidermal growth factor; ricin; tetanus; hexosaminidase;
KW Shiga-like toxin A; SLT-A; PH; ligand; insulin; nuclease.
XX
XX Vibrio cholera.
OS
XX
XX WO9109871-A.
PN
XX
XX 11-JUL-1991.
PD
XX
XX 21-DEC-1990; 90WO-US07619.
PF
XX
XX 14-JUN-1990; 90US-0538276.
PR 22-DEC-1989; 89US-0456095.
XX
XX (SERA-) SERAGEN INC.
PA
XX
XX Murphy JR;
PI
XX
XX WPI, 1991-222845/30.
DR N-PSDB; AAQ12712.
XX
XX Hybrid molecules for targeting chemical entity to cell - have
PT membrane translocating and cell binding-regions and used to
PT treat HIV infection, genetic enzyme-deficiency disorders etc.
XX
XX Disclosure; Fig 13(1-3); 59pp; English.
XX
XX Hybrid molecules are produced by covalently linking
CC (1) a portion (A) of the binding domain of a cell-binding ligand,
CC allowing binding of the mol. to an animal cell;
CC (2) a portion (B) of a translocation domain of a protein able to
CC translocate (C) across the cell cytoplasmic membrane, and
CC and (3) a portion (C) which is to be introduced into the cell.
CC (A) is derived from a steroid or polypeptide hormone, a single-chain
CC analogue of a monoclonal antibody able to bind an antigen expressed
CC on the cell surface, or a polypeptide toxin.
CC (B) is derived from a toxin (e.g. diphtheria toxin or Pseudomonas
CC exotoxin A).
CC (A) may be derived from insulin, interleukins 2, 3 or 6 or
CC epidermal growth factor.
CC Suitable enzymes in (C) include cholera toxin, ricin, tetanus toxin,
CC hexosaminidase A, protease, nuclease, Sfr-A, etc.
CC Specified examples are CT-A/DT-B/IL-2, SfrA/DT-B/IL-2,
CC ricin A/DT-B/IL-2, HIVP-BP/DT-B/IL-2 and the phenylalanine
CC hydroxylase-DT-B, or their biologically active mutants.
CC (CT-A= cholera toxin, DT-B= truncated diphtheria toxin,
CC SLTA= Shiga-like toxin A; HIVP-BP= HIV protease binding protein.
CC See also AAQ12710-12.
XX
XX Sequence 452 AA;
SQ
Query Match 11.3%; Score 213.5; DB 12; Length 452;
Best Local Similarity 26.6%; Pred. No. 1, 1e-12;

```

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

OY 104 RNLWRLSSRPSLMKSCPRFFLDYLAEGLSDFLDH-----QAVIKFELETHF 155
 DB 184 KKTWGVTFKTLKSLYKTHA---CYEYNHIFPLELEKCGFHEDNIPOLEDVSOFLQTCGF 240
 OY 156 SYYPVSGVAPHQYSLDODRYFPPIASVWRTLDKNSFLTPDLIHDLGHVPMILHPSFS 215
 DB 241 RLRPVAGLLSSRDFLGGLAFRVFHCQYIRHGSKPMWTPPEPICHELGHVPLFSDRSFA 300
 OY 216 EPPINMGRLEFTKVIKYOALPSSKKORIQTQSNLAIIVRCFMFTVESGLIENHGRKAYG 275
 DB 301 QFSOEIG-----LASLGAPEYIEKLAT-----IYWFVEFGLCCKGDSIKAYG 344

OY 276 AVLISSEPOLGHAFIDNVRVLPLELDQIIRLPNTSTPQETLSIRHFDLVE 328
 DB 345 AGLISSFGELOYCLSEKPKLPLELEKTAIONVTVEFPQPLVYVASEFNDAKE 397

RESULT 14
 AAW25788
 ID AAW25788 standard; Protein; 452 AA.
 AC AAW25788;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Human phenylalanine hydroxylase.
 XX
 KW Phenylalanine hydroxylase; human; hybrid protein; cell delivery;
 KW cell binding ligand; translocation domain; diphtheria toxin B';
 KW phenylketonuria; therapy.
 XX
 OS Homo sapiens.
 XX
 PN US5668255-A.
 PD 16-SEP-1997.
 XX
 PF 07-JUN-1984; 84US-0618199.
 XX
 PR 27-JUN-1991; 91US-0722484.
 PR 07-JUN-1984; 84US-0618199.
 PR 25-APR-1985; 85US-0726808.
 PR 07-JUN-1985; 85US-0742554.
 PR 22-DEC-1989; 89US-0456095.
 PR 14-JUN-1990; 90US-0538276.
 PR 04-AUG-1993; 93US-0102387.
 XX
 PA (SERA-) SERAGEN INC.
 XX
 PI Murphy JR;
 XX
 DR WPI; 1997-470103/43.
 DR N-PSDB; AAT91639.
 XX
 PT New hybrid molecules for delivery of agents to cells - comprise a
 PT binding domain of a cell binding ligand and a portion of a
 PT translocation domain of a protein
 XX
 PS Example 5; Fig 13A-C; 30pp; English.

XX This protein comprises for human phenylalanine hydroxylase (PH).
 CC A DNA molecule (see AAT91638) encoding PH was used to construct a
 CC PH-diphtheria toxin B' gene that was expressed in E. coli. The
 CC resulting hybrid protein can be used in the treatment of
 CC phenylketonuria. The active PH enzyme is targeted to, and
 CC incorporated into, the broad range of cells which native diphtheria
 CC normally attacks, achieving the widespread therapy that is needed
 CC for this inherited disorder. Claimed hybrid proteins comprise a
 CC translocation domain and a cell binding domain. They can be used
 CC for the delivery of agents (e.g. therapeutic genes, toxins,
 CC detectable labels) into cells. The use of a translocation

CC mechanism ensures that the hybrid will be effective in relatively
 CC low doses, since a high proportion of the substance of interest
 CC will be taken into the targeted cells. The hybrid molecules can be
 CC manufactured as a single hybrid recombinant protein, permitting
 CC reproducibility, consistency, and the precise control of
 CC composition.

XX
 SQ Sequence 452 AA;
 Query Match 11.3%; Score 213.5; DB 18; Length 452;
 Best Local Similarity 26.6%; Pred. No. 1,1e-12;
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

OY 104 RNLWRLSSRPSLMKSCPRFFLDYLAEGLSDFLDH-----QAVIKFELETHF 155
 DB 184 KKTWGVTFKTLKSLYKTHA---CYEYNHIFPLELEKCGFHEDNIPOLEDVSOFLQTCGF 240
 OY 156 SYYPVSGVAPHQYSLDODRYFPPIASVWRTLDKNSFLTPDLIHDLGHVPMILHPSFS 215
 DB 241 RLRPVAGLLSSRDFLGGLAFRVFHCQYIRHGSKPMWTPPEPICHELGHVPLFSDRSFA 300
 OY 216 EPPINMGRLEFTKVIKYOALPSSKKORIQTQSNLAIIVRCFMFTVESGLIENHGRKAYG 275
 DB 301 QFSOEIG-----LASLGAPEYIEKLAT-----IYWFVEFGLCCKGDSIKAYG 344

OY 276 AVLISSEPOLGHAFIDNVRVLPLELDQIIRLPNTSTPQETLSIRHFDLVE 328
 DB 345 AGLISSFGELOYCLSEKPKLPLELEKTAIONVTVEFPQPLVYVASEFNDAKE 397

RESULT 15
 AAY55893
 ID AAY55893 standard; Protein; 452 AA.
 AC AAY55893;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Human phenylalanine hydroxylase.
 XX
 KW Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;
 KW translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;
 KW shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;
 KW cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;
 KW adipocyte; cancer; virus; infection; antibody.
 XX
 OS Homo sapiens.
 XX
 PN US5965406-A.
 PD 12-OCT-1999.
 XX
 PF 07-JUN-1995; 95US-0488246.
 XX
 PR 04-AUG-1993; 93US-0102387.
 PR 07-JUN-1984; 84US-0618199.
 PR 27-JUN-1991; 91US-0722484.
 PR 25-APR-1985; 85US-0726808.
 PR 07-JUN-1985; 85US-0742554.
 PR 22-DEC-1989; 89US-0456095.
 PR 14-JUN-1990; 90US-0538276.
 XX
 PA (SERA-) SERAGEN INC.
 XX
 PI Murphy JR;
 XX
 DR WPI; 1999-632431/54.
 DR N-PSDB; AAZ30664.
 XX
 PT Recombinant DNA molecule encoding a three part hybrid protein used in
 PT the treatment of Aids and genetic deficiency diseases -
 XX
 PS Example 5; Fig 13; 31pp; English.

XX The invention relates to a recombinant DNA molecule encoding a hybrid
 CC protein comprising three parts: (a) the first part comprises a portion
 CC of the binding domain of a cell-binding polypeptide ligand allowing the
 CC hybrid protein to bind to an animal cell; (b) the second part comprises
 CC a portion of a translocation domain of a naturally occurring protein
 CC selected from diphtheria toxin, botulinum neurotoxin, ricin, cholera
 CC toxin, LT toxin, C3 toxin, shiga-like toxin, pertussis toxin and tetanus
 CC toxin, which translocate the third part of the across the cytoplasmic
 CC membrane into the cytosol of the cell; and (c) the third part comprises
 CC a polypeptide entity to be introduced into the cell, which is non-native
 CC to the naturally occurring protein of (b). This sequence represents the
 CC human phenylalanine hydroxylase for use in generating the hybrid of the
 CC invention. The hybrid molecule enables the use in generating of appropriate
 CC therapy to affected cells, allowing them to function properly and
 CC alleviate or cure the disease. The hybrid is especially used in treating
 CC genetic deficiency diseases, by delivering to affected cells an enzyme
 CC supplying the missing function, to supplementing cellular levels of a
 CC particular enzyme or a scarce precursor or cofactor, to directing toxins
 CC or other poisons to destroy particular cells (such as adipocytes, cancer
 CC cell, or virus infected cells), to counteracting viral infections, such as
 CC HIV, by introducing appropriate antibodies to viral proteins. It is also
 CC involved in the process of getting non-therapeutic substances such as
 CC detectable labels into cells.

XX Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 20; Length 452;

Best Local Similarity 26.6%; Pred. No. 1.1e-12;

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWYRLSSRFLSKSYCPFRFLDYLEAFGLLSDFLDH-----QAVIKFPELETHF 155
 DB 184 KKTGTGTVFKTLKSLYKTHA---CYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQCTCTGF 240
 QY 156 SYYPVSGFVAPHQVLSLLQDRYPIASVNRITLDKDNFSLTDPDLHDLGHVPLWLLHPSFS 215
 DB 241 RLRPVAGLLSSRDFLGGALPRVPHCTQYIRHSGSKPMYTPEDDICHLLGHVPLFSDRSPA 300
 QY 216 EFTINMGRFLTKVIEKVOALPSKKQRIQTLQSNLAIIVRCFWFTVESGLTENHGRKAYG 275
 DB 301 QFSQEIG-----LASLGAPDEVIEKLA-----IYWFTVEGLCKQGSIKAYG 344
 QY 276 AVLJSSSQELGHAFIDNVRVPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328
 DB 345 AGLSSSGELOYCLSEKPKLLPLELEKTAIQNTVTPEPQPLYYVAESFNDAKE 397

RESULT 16

AA78593

ID AA78593 standard; Protein; 452 AA.

XX AC AA78593;

XX 05-MAY-2000 (first entry)

XX Human phenylalanine hydroxylase protein sequence.

XX Phenylalanine hydroxylase; translocation domain; cell destruction;
 KW cell binding domain; genetic deficiency disease; cell targeting; cancer;
 KW adipocyte; enzyme delivery; anti-viral; HIV.

XX Homo sapiens.

XX US6022950-A.

XX 08-FEB-2000.

XX 07-JUN-1995; 95US-0479510.

XX 07-JUN-1994; 84US-0618199.

XX 27-JUN-1991; 91US-0722484.

XX 25-APR-1995; 85US-0726808.

PR 07-JUN-1985; 85US-0742554.
 PR 22-DEC-1989; 89US-0456095.
 PR 14-JUN-1990; 90US-0538276.
 PR 04-AUG-1993; 93US-0102387.
 PA (SERA-) SERAGEN INC.
 XX Murphy JR;
 XX WPI; 2000-160390/14.
 DR N-PSDB; AAZ90020.
 XX New two-part hybrid protein comprising a translocation domain and a
 PT cell-binding domain, for treating genetic deficiency diseases, cancer
 PT and HIV infections.
 XX Example 5; Fig 13; 32pp; English.

XX This sequence represents the phenylalanine hydroxylase protein sequence.
 CC The encoded protein can be included in the hybrid protein of the
 CC invention and used to destroy or modify the cell that the hybrid protein
 CC is targeted to. The hybrid protein comprises a first part which is a
 CC portion of the binding domain of a cell-binding ligand, effective to
 CC cause the hybrid molecule to bind to a cell of an animal. The second
 CC part comprises a portion of a translocation domain of a naturally
 CC occurring protein (e.g. the translocation domain of diphtheria toxin) the
 CC second part translocates the third part across the cytoplasmic membrane
 CC and into the cytosol of the cell. The third part comprises a chemical
 CC entity to be introduced into the cell, where each of the first and third
 CC part is non-native with respect to naturally occurring protein, and the
 CC covalent bond attaching the second and third part is cleavable. The
 CC phenylalanine hydroxylase protein can form part of the third portion of
 CC the hybrid protein. The cell binding domain binds to a specific cell and
 CC the translocation domain transfers the hybrid molecule across the cell
 CC membrane into the cytosol. The third part of the protein, linked to the
 CC translocation domain through a cleavable bond, can then carry out its
 CC function. The hybrid molecules are useful for treating genetic deficiency
 CC diseases by delivering to affected cells an enzyme supplying the missing
 CC function, to supplement cellular levels of a particular enzyme or a
 CC scarce precursor or cofactor, to direct toxins or other poisons to
 CC destroy particular cells (such as adipocytes, cancer cells, or
 CC virus infected cells), and to counteract viral infections such as HIV by
 CC introducing into appropriate cells antibodies to viral proteins.

XX Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 21; Length 452;

Best Local Similarity 26.6%; Pred. No. 1.1e-12;

Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWYRLSSRFLSKSYCPFRFLDYLEAFGLLSDFLDH-----QAVIKFPELETHF 155
 DB 184 KKTGTGTVFKTLKSLYKTHA---CYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQCTCTGF 240
 QY 156 SYYPVSGFVAPHQVLSLLQDRYPIASVNRITLDKDNFSLTDPDLHDLGHVPLWLLHPSFS 215
 DB 241 RLRPVAGLLSSRDFLGGALPRVPHCTQYIRHSGSKPMYTPEDDICHLLGHVPLFSDRSPA 300
 QY 216 EFTINMGRFLTKVIEKVOALPSKKQRIQTLQSNLAIIVRCFWFTVESGLTENHGRKAYG 275
 DB 301 QFSQEIG-----LASLGAPDEVIEKLA-----IYWFTVEGLCKQGSIKAYG 344
 QY 276 AVLJSSSQELGHAFIDNVRVPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328
 DB 345 AGLSSSGELOYCLSEKPKLLPLELEKTAIQNTVTPEPQPLYYVAESFNDAKE 397

RESULT 17

AA813327

ID AA813327 standard; Protein; 532 AA.

XX AA813327;

XX

DT 12-JAN-2001 (first entry)
 XX
 DE Caenorhabditis elegans Cod-5.
 XX
 KM Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;
 KM daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;
 KM PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes.
 XX
 OS Caenorhabditis elegans.
 XX
 PN W0200033068-A1.
 XX
 PD 08-JUN-2000.
 XX
 PF 02-DEC-1999; 99WO-US28529.
 XX
 PR 03-DEC-1998; 98US-0205658.
 XX
 PA (GENE) GEN HOSPITAL CORP.
 XX
 PI Ruvkun G, OGS S;
 XX
 DR WPI: 2000-423022/36.
 XX N-PSDB; AAA91624.
 PT Diagnosing and treating obesity and impaired glucose tolerance using
 XX modulators of daf-18 expression and/or activity -
 XX
 PS Disclosure; Fig 42; 402pp; English.
 XX
 CC The present sequence is Cod-5 from Caenorhabditis elegans. Cod-5
 CC is the aromatic amino acid hydroxylase that synthesizes serotonin from
 CC precursor L tryptophan. The cod-5 gene was knocked out to produce mutants
 CC completely lacking in serotonin. These mutants were found to have defects
 CC in metabolic control. A number of C. elegans proteins that have
 CC mammalian homologues acting in the insulin signalling pathway were also
 CC identified. The C. elegans age-1 gene encodes a homologue of the
 CC mammalian PI 3-kinase whilst daf-2 encodes a homologue of the mammalian
 CC insulin receptor. The C. elegans AKT kinase and PKB kinase act downstream
 CC of daf-2 and age-1, just as their mammalian homologues act downstream of
 CC insulin signalling. The C. elegans PTEN lipid phosphatase homologue,
 CC DAF-18, has been found to act upstream of AKT in the pathway. This
 CC discovery has enabled mammalian PTEN action to be mapped to the insulin
 CC signalling pathway. Conserved DAF motifs can be used to design probes to
 CC identify mammalian DAF homologues and thus to identify individuals with a
 CC predisposition towards the development of glucose intolerance conditions,
 CC such as obesity and diabetes.
 CC
 SQ Sequence 532 AA;
 XX
 Query Match 10.9%; Score 206.5; DB 21; Length 532;
 Best Local Similarity 25.2%; Pred. No. 7.1e-12;
 Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;
 XX
 QY 104 RNLW---YRLSSRFLSKSYCPFRFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155
 DB 251 RKTWGIYRKLK---ELHKHAKCKQFLDNFELLERHCGYSENNIPQLEIDICFLKAKG 307
 QY 156 SYYPVSGFVAPHOYLSDODRYFPRIASVMTLDKNSFLTPDLIHDLGHVPMILHPSFS 215
 DB 308 RVPVAGYLSARDPLAGLAVRVFCTQYVRHMDPFTYRPEDVHELMGHMALFADPDFA 367
 QY 216 EFTINMGRLEFTVKIEKVALPSKKORIOTQSNLIAIVRCFWFTVSSGI-----265
 DB 368 QFSQETG-----LASTASBEDDKLATL-----YFSSISFGLSSDDADSPVK 411
 QY 266 ---ENHEGRKAVGAVLSPOELGHAFIDNVRVPLEDITLTPNTSTPQETLPSIRH 322
 DB 412 ENGSNHERKAVGAGLSAGLQAVAGSGRTIRPPRDVVEQGLITTFQSAFYTRN 471
 QY 323 FDELVELTSGKLEMLDQGLLESIPLYNOEKYLSGEVL 360
 DB 472 FEBAQO---KLRFMTNNMKRPPIVRYN--PYTESVEVL 504

RESULT 18
 AAU30703
 ID AAU30703 standard; Protein; 491 AA.
 XX
 AC AAU30703;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #1194.
 XX
 KM Human; vaccination; gene therapy; nutritional supplement;
 KM stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KM immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN W0200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 XX 26-JAN-2001; 2001US-0770160.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 XX vaccination, testing and therapy -
 XX
 PS Claim 20; Page 332; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemia. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 CC
 SQ Sequence 491 AA;
 XX
 Query Match 10.6%; Score 199.5; DB 22; Length 491;
 Best Local Similarity 28.4%; Pred. No. 3.2e-11;
 Matches 58; Conservative 36; Mismatches 83; Indels 27; Gaps 5;
 XX
 QY 104 RNLWYRLSSRFLSKSYCPFRFLDYLEAFGLSDFLDH-----QAVIKFPELETHF 155
 DB 218 KKTWGIYRKLKSLVYTHA---CYEYNNHIFLEKYKASHEDNIPQLEIDVSGFLQCTGF 274
 QY 156 SYYPVSGFVAPHOYLSDODRYFPRIASVMTLDKNSFLTPDLIHDLGHVPMILHPSFS 215
 DB 275 RLRFVAGLSRDLGLDAFRVHCTQYIRHSGSKMTYRPEDVHELMGHMALFADPDFA 334
 QY 216 EFTINMGRLEFTVKIEKVALPSKKORIOTQSNLIAIVRCFWFTVSSGI-----275
 DB 335 QFSQETG-----LASTGAPDESIEKL-----APIYFTVREGLCQGDSTKAYG 378

QY 263 GLIENHEGR-KAYGAVLISSPQELGHA 288

303 QEIQ-----LASLGAPDDYIEKLST-----IFWFTVEYGVCRQEGELKAYGAGL 346

Qy 263 GLIENHEGR-KAYGAVLISSPQELGHA 288

ABP06168
ID ABP06168 standard; Protein; 104 AA.
XX AC ABP06168;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX protein sequence SEQ ID NO:12318.
XX DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX OS Homo sapiens.
XX OS WO200192523-A2.
XX PN 06-DEC-2001.
XX PD 29-MAY-2001; 2001WO-US10836.
XX PF 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PR (CUFA-) CURAGEN CORP.
XX PA Shinkets RA, Leach MD;
XX PI WPI; 2002-106308/14.
XX DR N-PSDB; ASN21920.
XX DR Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX Disclosure; SEQ ID 12318; 1037pp; English.
XX The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, Graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 104 AA;
XX SQ

Query Match 5.8%; Score 109.5; DB 23; Length 104;
Best Local Similarity 28.6%; Pred. No. 0.0041;
Matches 26; Conservative 20; Mismatches 44; Indels 1; Gaps 1

QY 125 FF-----LDYLEAFGLSDFLDHOAVIKPFEELETHFSYYPVSGFVAHQVLSLDORFFP 179
 Db 943 SHSDVVOGQMKALGVLTVIARQ---KGFS-----SRLLPDEFLLRRQGFALFDPDE--P 992
 QY 180 TASWRTLDKDNFSLTPDLIHL-----LGHVPMILHPSFSFPIINMGLPTKYIEKYQ 233
 Db 993 V-----EMTKDNCRLL--LFLRLKMGMLGKTKVFLRYNDEF--LAKRLYELOVKRY- 1040
 QY 234 ALPSKKQRIQTLSNLIAYRCFWFYVESGLI-----ENHGRKAYGAVLISPQEL 285
 Db 1041 -----IKVQSWMRALLARKR-----VKGKVFELGKGGPEHND-----VAASKIQKA 1082
 QY 286 GHAFIDNVRVLELELDQIIRL-----PFNTSTPQETLFSI-----RHFDVLYLT 330
 Db 1083 FFGFDRVRLPPLVNEKSGQLENTADFIREFAKWREKSIQVLLHYRAARFQDFVNL 1142
 QY 331 SKLEWMLDQGLLESIPLYNQ 350
 Db 1143 -----QGVHIYNQ 1150

RESULT 25

ABR35685
 ID ABR35685 standard; Protein; 1782 AA.

XX ABR35685;

AC 26-MAR-2002 (first entry)

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 23847.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL09788.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Disclosure; SEQ ID NO 23847; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutic and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

XX sequences (AB16176-AB16175) and the encoded proteins

XX (AB16176-AB16175).

XX The sequence data for this patent did not form part of the printed

Query Match 5.5%, Score 103.5, DB 22; Length 1782;

Best Local Similarity 25.0%; Pred. No. 0.99;
 Matches 73; Conservative 32; Mismatches 96; Indels 91; Gaps 15;
 QY 67 ILLEF-----FKNLLFYHLISLSTKNOREGCSIDMAV-----VSPFENRLWYR 109
 Db 574 ILLEPQAIYVNHYSYRNLFFSPKELNFFSRASASANINVRQWAGEPKQAVAIY- 632
 QY 110 LLSRFLMKSCGRRPFLDYLEAGLS-----DFLDHOAVIKFELETH-----PSYRP 159
 Db 633 -----GKSCPKF--STAFYAVYNNKCSFIDELIKALPASIKHHLLFTIYH 681
 QY 160 VSGFVAP-----HOYLSLDORFFPIS-----VMTLKDKNFSLTPDLIHL 203
 Db 682 VSCCKRPQDLOPSVETPIGYTWLPLEDGLKFKGEPNLPMVWSPPENVSFTPEVNH--L 739
 QY 204 GHVPMILHPSFSFPIINMGLFTKYIEKYQALPSKKQRIQTLSNLI-A-IVRCFWFYVES 262
 Db 740 FGIKML-----DNHRAVFSINVEAVTA-----IHTLDSFLDRPFLICEYLDTEN 783
 QY 263 GLIENHGRKAYGAVLISPQELGHAFID-----NVRVLELELDQIIRL 306
 Db 784 --IPSHIGEN-----NIETELKKCLDIEYANRREPLVRLPLVDKLEL 826

RESULT 26

ABR35686
 ID ABR35686 standard; Protein; 563 AA.

XX ABR35686;

AC 24-JUL-2002 (first entry)

DT 24-JUL-2002 (first entry)

DE Fungal ZBC protein sequence #112.

XX Secondary metabolite; fungus; ZBC gene; zinc binuclear cluster protein;

XX antibacterial; beta-lactam; anti-hypercholesterolaemic; lovastatin;

XX mevastatin; immunosuppressant; cyclosporin A; ergot alkaloid; ergotamine;

XX angiotensin inhibitor; ovalicin; glucan synthase inhibitor; gliotoxin;

XX fungal toxin; cell surface receptor; plant growth regulator; pigment;

XX insecticide; antineoplastic.

XX Unidentified.

XX WO200224865-A2.

XX 28-MAR-2002.

XX 19-SEP-2001; 2001WO-US29288.

XX 19-SEP-2000; 2000US-233564P.

XX (MICR-) MICROBIA INC.

XX Holtzman D, Madden K, Maxon M, Sherman A;

XX WPI; 2002-352005/38.

XX N-PSDB; ABR79875.

XX New method for improving the production of a secondary metabolite e.g.

XX antineoplastic agent, ergot alkaloid from a fungus involves modulation

XX of the expression of at least one zinc binuclear cluster protein gene

XX -

XX Disclosure; SEQ ID 236; 49pp + sequence listing; English.

XX The invention relates to improving the production of a secondary

XX metabolite by a fungus. This involves modulating the expression of at

XX least one ZBC (zinc binuclear cluster protein) gene in a manner to

XX improve the yield of the secondary metabolite. Methods of the invention

XX may be used for improving the production of a secondary metabolite e.g.

XX antibacterial (such as beta-lactam), an anti-hypercholesterolaemic

XX (such as lovastatin), an immunosuppressant (such as cyclosporin A),

XX an ergot alkaloid (such as ergotamine), an angiotensin inhibitor (such

XX as lovastatin).

CC as ovalicin), a glucan synthase inhibitor, gliotoxin family of compounds,
CC a fungal toxin, a modulator of cell surface receptor signaling; a plant
CC growth regulator, a pigment, an insecticide, or an antineoplastic
CC compound. The method results in an increase in fermenter run-time, a
CC decrease in the size of the fermenter required for the production of
CC equivalent amounts of the secondary metabolite, or a decrease in the
CC biomass required for the production, which translates into decreased
CC waste that must be handled in downstream processing. The sequences given
CC in records ABP35575-ABP35702 represent ZBC proteins.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Seq Sequence 563 AA;

Query Match 5.4%; Score 102; DB 23; Length 563;

Best Local Similarity 21.1%; Pred. No. 0.27;

Matches 80; Conservative 56; Mismatches 142; Indels 102; Gaps 16;

QY 7 TLDPKYILKIALKROSLSIFFQNSQSORAYSTPYSYRILOKENKEQALARKHCIS 66

Db 185 TLLPTIIATIQSLDPLVDVILNFYNSAGI-----TPLESSRLINELKNEISEQYKHLCLP 239

QY 67 ILEFFKNLLFVHLLSLSKNOREGCGTDM-----AVVSTPPFNENLWYRLLSRFSLWKS 120

Db 240 DKEIIQMLLRAYATKFRIRGVNTDLCSRHVSTLVPLF--QVTEKIGKNTSLWFA 297

QY 121 YCPREFDYLEAFLGSLDSFLDHQAVIKPFELETHESYYPVSGFVAPHOYLSLLQDRYFPI 180

Db 298 LC-----EIDGL-----EVLKX-----RPPIQDITVGRUKP----- 325

QY 181 ASVVRTLDKONFSLITPDLIHLLGHVPMLLHPSSEFPFNNGRFLTQVIEKVOALPSKKQ 240

Db 326 ---LRCPFNDDISYN---PHLLGR---LLDCGVSIYKSVHSLTVSKFDIKLESYESQLS 376

QY 241 RI-----OTLQSNLIATVRCFW-----FTVESGLIENHEGRKAYGAV 277

Db 377 LILVDIEAKFYDPSNEDIQFRIFLKMVFMTARVNLVQCFTILDGILEDEE-----T 429

QY 278 LISSPQELGHAFIDNVRVLPLELDQIIRLPFNFTSPQETLFSIRHFDLVELTSLKLEWML 337

Db 430 IIGN---LGESCIQCVRLD---ISQITIL-----EKRGLWLVALLEIHALMLAAFCR 476

QY 338 DOGL-----LESIPLYNQEK 352

Db 477 DKGFEVPSDLGDLITLVQER 496

RESULT 27

AAB94081

ID AAB94081 standard; Protein; 456 AA.

XX

AC AAB94081;

XX

DT 26-JUN-2001 (first entry)

XX

DE Human protein sequence SEQ ID NO:14279.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX

OS Homo sapiens.

XX

PN EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-0116126.

XX

PR 29-JUL-1999; 99JP-0248036.

XX

PR 27-AUG-1999; 99JP-0300253.

XX

PR 11-JAN-2000; 2000JP-0118776.

XX

PR 02-MAY-2000; 2000JP-0183767.

XX

PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 14279; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides, and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

Seq Sequence 456 AA;

Query Match 5.3%; Score 101; DB 22; Length 456;

Best Local Similarity 23.3%; Pred. No. 0.25; Mismatches 98; Gaps 20;

Matches 81; Conservative 41; Indels 128; Gaps 20;

QY 63 KCISILEFF-KNLLFVHLLSLSKNOREGCGTDMVVSTPPFNENLWYRLLSRFSLWKS 121

Db 70 KCLVEHFHFGFNEKRVNLAAQRR-----SVPLL-RAISYHLVQKPFSLTKD- 118

QY 122 CRRFLDYLEAFGLLS-----DFLDHQ-----AVIKPFBELETH 154

Db 119 ---VLLDVAYAYGKLSFPHOTQVSQRATDLSLMPESLTSGEVAHCAKSFALLKLSL--- 172

QY 155 FSYVPVSGFVAPHQYLSLLQDRYFP-IASVVRTLDKONFSLTPD-----LIHDLG-- 204

Db 173 ----PLFGAFAOH-VLNRAQDITLPHLCVSLIAFARLNFH--PDQEDQFFSLVHEKLGSE 225

QY 205 -----HVPW-----LLHPSFSEFFINMGRLFTKVIKQVQALPSKK 239

Db 226 LPGLEPALQVDLVWALCVLQQAERAEALQVLPFP-----HIQFLGGKS 269

QY 240 QRIOTLQSNLIATVRCFWFTVESGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLE 299

Db 270 QKDQNTFQKLHI-----NATALLEYPE---YSGPLLPASAVAPGPSALDR-KVTPLQ 318

QY 300 LDQIIRLPFNFTSTPQETLFSIRHFDPE--LVELTSLKLEWMLD-QQLLES 344

Db 319 KEL-----QETLKGLLGSADKGSLEVATQYQWGLDAEVLDD 355

RESULT 28

AAB59689

ID AAB59689 standard; Protein; 2042 AA.

XX


```

Db 233 KVEGSMQKLEN--VLNRASNTADTLFOEVLGRKDK--ADSTRNALNVLFQFK-FLFNL 286
Qy 79 LLSLSKQREGCSTDMVAVSTPFENRWYLLSSRF-----SLWKSYPFRFFDYLEAF 133
Db 287 FLNKRNIQKG--DYDVVIND-----YEKAKSLFGKTEVQVFKY-----YAEVE 329
Qy 134 GLLSDFLDHQAIVKFFELTHESYVPVSGFV-----APHQY-LSLLO 174
Db 330 AGIEDL--RELLKKL-LETPSLDQKRYIRYLSDLHAPCDPAWQCIGAQHWLKLQ 386
Qy 175 D-----RYFPASVMRTLDKDNESLTPDLHDLHG----- 206
Db 387 DCKEGHMKSLKGNPGHSPM-----LQDN-----DARPSVLGHLSQTASLKRGSFQS 435
Qy 207 -----PWLH-----PSFSEFFIM--GRLFKTVIEK---VQALPSK 238
Db 436 GRDTRWRYKTHRAVFAVEKLTKLIVLSQPNFKLWISYVNGSLFSETAEKSGQIERSKNV 495
Qy 239 KOR-----IQLOSNLIAVR--CFWFTVESGLIENHGRKAYGAVLIS---SPQELG 286
Db 496 RORQNDFFKMIQEVHSLVKLIRGALLPFLREG-----DGRQ-YGGEVQVQALSGQWLA 549
Qy 287 HAFIDNVRV-----LFLDQIIRLPENTSTPQETLFSIR-----HFDELV 327
Db 550 HV-IQTIRLVESLTALEIPNDMLQII-----QDLILDLRIHCIMVTLQHTAEIK 599
Qy 328 ELTSKLEWMLDQGLLESIPLNQEKVLSGFVL 360
Db 600 RLAEKEDWIDNEGLTSLPCQEQSIVHLSQSL 632

```

RESULT 30

AAW19604

ID AAW19604 standard; Protein; 1024 AA.

AC AAW19604;

DT 21-AUG-1997 (first entry)

DE Mycoplasma genitalium 116 kDa protein MG075 useful in vaccine.

KW Mycoplasma; immunogen; vaccine; diagnosis; pneumonia; inflammation.

OS Mycoplasma genitalium.

PN WO9721727-A1.

PD 19-JUN-1997.

PF 13-DEC-1996; 96WO-AU00803.

PR 13-DEC-1995; 95AU-0007127.

PA (UYME) UNIV MELBOURNE.

PI Browning GF, Duffy MF, Walker ID, Whithear KG;

DR WPI; 1997-332722/30.

PT New immunogenic polypeptide(s) from Mycoplasma species - useful in
 vaccines and for diagnosis of Mycoplasma infection

PS Claim 19; Page 85-89; 110pp; English.

CC Isolated or recombinant immunogenic polypeptides from Mycoplasma
 genitalium have mol.wt. of 16 kDa (AAW19603) (MG074) and 116 kDa
 (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins
 (see also AAW19601-02) obtd. from Mycoplasma pneumoniae. A genomic
 DNA sequence of M. genitalium contains contiguous open reading
 frames that code for the 2 polypeptides. Mycoplasma 16 or 116 kDa
 proteins, or immunogenic fragments that include a T or B cell
 epitope, can be used in vaccines for prevention and treatment of

CC Mycoplasma infections, partic. in humans. They can also be used
 CC diagnostically to detect Mycoplasma, or to raise antibodies useful
 CC in immunoassays.

SQ Sequence 1024 AA;

Query Match 5.2%; Score 98.5; DB 18; Length 1024;

Best Local Similarity 19.9%; Pred No. 1.4;

Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;

Qy 29 QNSQISQIRASTSYRRILOKENKEX---QALARHKCISILEFFKNLLFVHLLSLKN 85

Db 564 QQTDSLKNLFSV---IGDILSETNVNKTITHAVNNELLSLVETASTLKIKHL----- 613

Qy 86 QREGCSTDMAVSTPFENRWYLLSSRFSLWKSYPFRFFDYLEAFGLLSDFLDHQA 145

Db 614 -----NVQYKVLVDKFKNSFTK-----ELLNFFDTKDITPT 647

Qy 146 IK--FFELETH-----FSYYPVSGFVAPHOVLISLQDRYFPFIASVMTLKD 192

Db 648 IKKVLFESENYKTLRKKNYENEGPGYHWAFTIVPCTFNSAENTFYSAI-----DKT 698

Qy 193 SLTPDLIHDLIHVPWLLHPSFSEFFINNGRLFT-----KVIEKVALPSKKORIQT 244

Db 699 KSIRDLPADML-FGKSLSVNDSDSFIKINGSFTLITKNDNLNLLPNYHSLITKNVGYQI 757

Qy 245 LOSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISSPQELG----- 286

Db 758 VVNVFHIDARLLTAELQNTVFSNPK-----PVIKSPVELSKSLFEVWKTIFENS 810

Qy 287 ----HAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRHFD 324

Db 811 LKKEYTFEDNLKFFPFKADGSSLEFDSLKDQVRVIFPAFVD 852

RESULT 31

ID AAE11093 standard; peptide; 56 AA.

AC AAE11093;

DT 18-DEC-2001 (first entry)

DE phhAB fusion peptide #4.

KW Phenylalanine hydroxylase; PAH; phhA; 4a-carbinolamine dehydratase; phhB;
 KW aromatic aminotransferase; phhC; milk protein; animal protein; casein;
 KW proteinaceous food product; globulin; whey protein; phenylketonuria;
 KW PKU; inherited metabolic disorder; impaired brain function; nootropic;
 KW cell therapy; phhAB fusion peptide.

OS Unidentified.

PN WO200168822-A2.

PD 20-SEP-2001.

PF 14-MAR-2001; 2001WO-DK00172.

PR 14-MAR-2000; 2000US-0525116.

PA (NITL-) NILAB APS.

PI Johnsen M, Ravn P, Madsen SM, Vrang A, Israelsen H, Bredmose L;

PI Arnau J, Jensen SH, Gjetting T, Nielsen E;

XX WPI; 2001-590055/66.

PT Novel recombinant cells comprising a nucleic acid encoding a gene
 product having phenylalanine hydroxylase activity, that is derived from
 a prokaryotic organism, is useful for treating phenylketonuria in
 mammals

XX

PS Disclosure, Page 19, 91pp; English.

XX The patent discloses novel cells comprising a nucleic acid encoding

CC a gene product having phenylalanine hydroxylase (PAH) activity such

CC as phenylalanine hydroxylase (paha), 4a-carbinolamine dehydratase

CC (pnhb) and aromatic aminotransferase (pnhc), which are derived

CC from a prokaryotic organism. The patent also relates to fusion

CC proteins comprising a protein enhancing and/or stabilizing the

CC PAH activity in addition to PAH activity. The cells are useful

CC for preparing a proteinaceous food product (animal protein such

CC as a milk protein derived from casein, globulin or a whey protein)

CC having reduced content of phenylalanine. The method involves contacting

CC the food product starting material with the cells or fusion proteins

CC such that at least part of the phenylalanine content of the starting

CC material is converted into compounds that do not cause phenylketonuria

CC (PKU) by the enzymatically active product. PAH enzyme is useful for

CC manufacturing a medication for treating PKU, which is an inherited

CC metabolic disorder resulting in an accumulation in the body of

CC L-phenylalanine and metabolites that can cause impaired brain

CC function. The present sequence is a peptide sequence of pnhab

XX fusion protein.

XX Sequence 56 AA;

Query Match 5.2%; Score 98; DB 22; Length 56;

Best Local Similarity 35.7%; Pred. No. 0.024; Indels 0; Gaps 0;

Matches 15; Conservative 13; Mismatches 14;

OY 250 IAIYRCFWFVSGLIENHGRKAYGAVLISPOELGHAFFID 291

DB 1 VFLARLYWMTIEFGLVETDQKRIYGGGLISPKETVYSLSD 42

RESULT 32

ABB97206

ID ABB97206 standard; Protein; 689 AA.

XX ABB97206;

XX 28-JUN-2002 (first entry)

DT

DE Novel human protein SEQ ID NO: 474.

XX Human; anti-infective; anti-inflammatory; immunomodulator;

XX anti-infective; cerebroprotective; cytoprotective; rheumatic; gene therapy;

KW neuroprotective; antiparkinsonian; protein therapy; EST;

KW expressed sequence tag.

XX Homo sapiens.

OS

XX WO200222660-A2.

PN

XX 21-MAR-2002.

PD

XX 10-SEP-2001; 2001WO-US26015.

PF

XX 11-SEP-2000; 2000US-0659671.

PR

XX (HYSE-) HYSEQ INC.

PA

XX Tang YT, Liu C, Zhou P, Aouni V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

PI

XX WPI: 2002-292408/33.

DR

XX N-PSDB; ABRJ32392.

DR

XX An isolated polynucleotide for treating diseases associated with its

PT encoded polypeptide such as cancer and multiple sclerosis -

PT

XX Example 2; SEQ ID NO 474; 509pp; English.

XX The present invention provides the protein and coding sequences of 444

CC

CC novel human proteins. These were isolated from expressed sequences tags

CC (ESTs). They can be used to stimulate cell growth, to regulate

CC hematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth

CC e.g. in burn treatment, to regulate the immune system e.g. to treat

CC multiple sclerosis, to regulate activity or inhibit e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat

CC stroke and cancer, to screen for drugs, to treat inflammatory conditions

CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.

CC Parkinson's disease. The present sequence is a protein of the invention.

XX

XX Sequence 689 AA;

Query Match 5.2%; Score 98; DB 23; Length 689;

Best Local Similarity 22.2%; Pred. No. 0.9; Indels 114; Gaps 17;

Matches 74; Conservative 38; Mismatches 107;

OY 19 KLNQSLPLFONS-----OSLQRAYSTPYSYRITLQENKQALARRHKCISILEFPKN 73

DB 188 KAAGSLPLFNDHCLFTMLQELDKTPGSLHGRICIOALIDPKKAT-----AN 237

OY 74 L-LFVHLTLSSKQREGCSTDW-AVSTPFENRMWRLSRSFSLW-----KSY 121

DB 238 LKQFLBELNRHGRPAKCLTIWMLGQGFAN-----LLEGKWLGLMPLVLGKSL 290

OY 122 CRRFELDYLEAFGLSDPFDHQAIVKPELETHPSYVPSGVAPHOYLSDDRYFPFA 181

DB 291 SP-FAITYIDRLIM-----HNLTKGF-----GMIGKDPFPLDYAYMP-- 330

OY 182 SWRTLDKDNFSILTPILHDLIGHVPM-----LHPSFSFPFIMGRLETKYI 229

DB 331 -----NNSLTPSLQQLQLYRLKVLAFPAKDPSTLHTYPSFL----- 370

OY 230 EKVQALPS-----KKORIOTLQSLIA--IVRCFTV-----ESGLNHEGRAYGA 276

DB 371 --SRATPSCPPEMKKELSLSTECCTVDPCLASVWRQLYPRHQSLSLDEH----- 420

OY 277 VLISS-----POELGHAFFIDNREVLPELDQIIR 305

DB 421 -LLSSWEQIPKVKVQKSLQETIOSLKLITNOELIR 452

RESULT 33

AAV35368

ID AAV35368 standard; Protein; 439 AA.

XX AAV35368;

XX 13-SEP-1999 (first entry)

DT

DE C. pneumoniae protein involved in metabolism of nucleic acids.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

OS

XX WO9927105-A2.

PN

XX 03-JUN-1999.

PD

XX 20-NOV-1998; 98WO-IB01890.

PF

XX 04-NOV-1998; 98US-0107078.

PR

XX 21-NOV-1997; 97FR-0014673.

PR

XX (GEST) GENSET.

PA

XX Griffais R;

PI

XX WPI: 1999-357842/30.

DR

XX Genome sequence of Chlamydia pneumoniae

PT


```
XX Page 1164; Disclosure; 1912pp; English.
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
XX frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
XX C. pneumoniae causes respiratory disease such as pneumonia and
XX bronchitis and is thought to be a contributing factor in heart
XX disease, sarcoidosis, sinusitis, purulent otitis media, erythema
XX nodosum or pharyngitis. The polypeptides encoded by the open reading
XX frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
XX immunogenic compositions as vaccines. Vectors containing C. pneumoniae
XX nucleotide sequences can also be used as immunogenic compositions,
XX especially where the vector directs the expression of a neutralising
XX epitope of C. pneumoniae.
XX
XX SQ Sequence 439 AA;
XX
XX Query Match 5.1%; Score 97; DB 20; Length 439;
XX Best Local Similarity 22.1%; Pred. No. 0.59; Indels 104; Gaps 19;
XX Matches 77; Conservative 51; Mismatches 117;
XX
XX QY 46 RIIQKENKEQALARKHCISIL--EFFKNLLF---VHLLSLKXNREGSGTDM----- 95
XX DB 138 RLIIHKNPALHQSOLVLIHTNYLQKDLKXNVLFQEPHLLAVRYNVTSKRTSLVDKLLA 197
XX
XX QY 96 ----VSTPTFFNR-----NLWY-RLLSSRFSLMKSCYPRFFLDYLEA-----FG 134
XX DB 198 SYTOPISYSYFSRVERLEQISLWHQOIVNSLLEIPK----QVFLDQLTAHISGFKKQPPS 253
XX
XX QY 135 LLSDFLDHQAIVKFFLETH---FSYYPVS-QFVAPHOYLSLLQDRVFFPIAS----- 182
XX DB 254 ILDDL--HHFVDLLYTSEHSSFSFPAETENFKRLA-----RYKPCAAFTVLENNMS 306
XX
XX QY 183 -VMRTLDKNP----SLTPDLIHDL-LGHVPMLLHPSEFFINMGRLETKVIEKVOAL 235
XX DB 307 WVERTLEPCNLDRIFNTLLVDLOEYLKQNYTPW-LSPDESVF-----ALEKLLSS 355
XX
XX QY 236 PSKQRTQTLQSNLLAIVCFWTFVSGLIENHGRKAYGAVLISSPOELGH---AFIDN 292
XX DB 356 SEAPVVOAL-----REGYQLVLIDEPQDTDKQWSIFSN 390
XX
XX QY 293 VRVPLELDQIIRLPNTSTPQTLSIHHPDELVELTSKLEWMLDQGL 341
XX DB 391 LFISPKFTGSL----FLIGDPKOSIYEWKSAADLPYLTAKSFSFSEDKQL 435
XX
XX RESULT 34
XX AAG31679
XX ID AAG31679 standard; Protein; 920 AA.
XX
XX AC AAG31679;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38084.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137724.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138640.
XX 10-JUN-1999; 99US-0138647.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 21-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 28-JUN-1999; 99US-0140891.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141642.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
```


DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 38083.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0125548.
XX 23-MAR-1999; 99US-0125786.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131149.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137504.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149358.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149317.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	13-OCT-1999;	99US-0159329.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144864.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144814.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0145086.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160741.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161320.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161392.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161593.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			

Query Match

5.0%; Score 95; DB 21; Length 970;

Best Local Similarity 21.7%, Pred. No. 3;
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

```

QY 9 DEPKYIKALKRQSLSPFONSQSLRA-----YSTSYSTRITLQENKEKQALR 61
DB 421 NPAIVOKVVDLSPSNFRIEMBSQFSDTKABPMYINVAISLEKITSS----- 469
QY 62 HKCISLIEFFKNLLFVHL-----LSLKNORBCSCSDMAVSTPPNNMYR 109
DB 470 ---TIQEWVQAPVHLHLPAVNFIFPTDLSL-KDADDEKTEVPVLRKTPF--SRLMYK 522
QY 110 LLSRPSLWKSYP-----CF-----RFLDYLEAF----- 133
DB 523 -PDTMFSKPKAVKMDPNCPLAVSPDAVLTIDFTLMDYNEVAYVAQVAGLYGVS 581
QY 134 -----GLLSDP---LDHQAVIKFELETFPSYYPVSGFVA 165
DB 582 LSDNGPELTLGYNHKLRILETYVGKIANEFVPRFAVIK---ETVTKKEYQVYKFRQ 637
QY 166 P-HQ---YLSL-LQDRYPPIASVMTLDKONFSLTPDLIDHLGH-----VPMLH 211
DB 638 PYHQAMYCSLLDQOTW-----WTEBL--DVLSHLEADVAKFVPMILS 681
QY 212 PSFSEFFI-----NMGRFTFYIEKV-----QALPSKKORIOTLOSNIATVRC 255
DB 682 RTIECTINGNVENNEAESMKHIEDVLFNDPKICRPFPS-----QHLTKRVKL--- 733
QY 256 FMFTVESGLIENHGRKAVGAVLISSPOELGHAFIDNVRVLPLELDQIIRLP-FNTSTPQ 314
DB 734 -----GEGMKKFFYHODGSPSDENSALVHYIVQHRDPSFMKIKLQFGLVAKQ 781
QY 315 ETLFSIRHFDLVELTSLKLEWMLDGL 341
DB 782 ATFHQDRIVEQLGYITA-LAQRNDGI 807

RESULT 37
AA055671
ID AA055671 standard; Protein; 293 AA.
XX
XX AA055671;
XX
XX 24-JUN-1998 (first entry)
XX
XX H. pylori ORF hp5p15641_5211687_c2_29 cellular protein.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX identification; binding compound; bactericidal; life cycle; activator;
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
XX
XX WO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
XX 02-APR-1996; 96US-0758731.
XX 25-OCT-1996; 96US-0736905.
XX 28-OCT-1996; 96US-0738859.
XX
XX (ASTR ) ASTRA AB.
XX
XX Alm RA, Smith D;
XX
XX WPI, 1997-503122/46.
XX
XX N-PSDB; AAV25080.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT

```

PT infection and for diagnosis of H. pylori infection

PS Claims 14,93; Pages 878-879; 1145pp; English.

XX This is the sequence of a Helicobacter pylori cellular protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds. The
 CC infection as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.

Sequence 293 AA;

Query Match 5.0%; Score 94; DB 18; Length 293;

Best Local Similarity 21.0%, Pred. No. 0.67; 67; Indels 38; Gaps 6;

Matches 38; Conservative 38; Mismatches 38; Indels 38; Gaps 6;

```

QY 179 PIASVMTLDKDNFSLTPDLIDHLGHVP-----W-----LIH 211
DB 93 PIAMTAQO-BEDVLKLSVLEIKDLTGKAPGVAPWMEFSNITNELLKRGFYDHSIMH 151
QY 212 PSFSEFFINMGRLFTFYIEKVQALPSKKORIOTLOSNIATVRCFMFTVESGLIENHGR 271
DB 152 NDFPYFVRVQDSWKIDSLAKMKKPLIRGVETINLVEIPANWYLDLPRMFFKSP 211
QY 272 KAVGAVLISSPOELGHAFIDNVRVLPLELDQIIRLP-FNTSTPQETLFS---IRHPD 324
DB 212 NSFGRV---SPRDIQGMWIDQFDWYRVEMDYAVFSMTIHDPVARSAPVLLMHKIEIHIN 268
QY 325 E 325
DB 269 Q 269

RESULT 38
AA082418
ID AA082418 standard; Protein; 528 AA.
XX
XX AA082418;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO.1930.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000MO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX

```


QY 209 LHP5--FSEFFINMGRLEFTKVEIKVQALPSKQRIQTLSNL--IAIVRCFWFTVESGL 264
 Db 297 SIHPSRQFDHYESN-----QVAKLLLL-----RINSTDDEHFEQIEDAIVSLVQ-L 344
 QY 265 IENHEGRK-----AYGAV-----LISPP-----OELGHAFID--- 291
 Db 345 AKVFSKQKRVSVYMMPSLYRLNLLITVGIITKVPTVIRKLISGGLLYLDSDNDKVFVQL 404
 QY 292 --NVRVLPLELDQIIRLEPNTSTPQETLFSRHPDELVELTSKL----- 333
 Db 405 LTNKISPLMKSQYNNVLRNWEYDKVELEFNFDPQVETLEQIKKRIISNDITNLQJLSK 464
 QY 334 -----EWMUD-----QGLLESPLVYNCKEYISGEVUC 361
 Db 465 TPLSIKIMVAMWYLSHLCSGLISV--NRIVLTKIRKIFC 502

RESULT 40
 ABB55057
 ID ABB55057 standard; Protein; 307 AA.

XX AC ABB55057;
 XX 16-MAY-2002 (first entry)
 XX DE Lactococcus lactis protein prnd.
 XX KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX OS Lactococcus lactis IL1403.
 XX PN FR2807446-A1.
 XX PD 12-OCT-2001.
 XX PF 11-APR-2000; 2000FR-0004630.
 XX PR 11-APR-2000; 2000FR-0004630.
 XX PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX PI Bolocrine A, Sorokine A, Renault P, Ehrlich SD.
 XX DR WPI; 2002-043418/06.
 XX PT New nucleotide sequence useful in the identification or Lactococcus
 XX PT lactis and related species -
 XX PS Claim 6; SEQ ID No 1759; 2504bp; French.
 XX CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO2001/77334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 307 AA:

Query Match 4.9%; Score 93; DB 23; Length 307;
 Best Local Similarity 20.9%; Pred. No. 0.9; 93; Indels 122; Gaps 16;
 Matches 67; Conservative 38; Mismatches 93;

QY 15 KIAI--KLRQST---SLPFONSQISQIRAVSTPYVYRIILQKE-----NKEQALARRKC 64
 Db 5 KVTLDKKIRSVWMSRMLQGSWMYERNQNGMAYSLIPALKKLYSGEBAKALKRH-- 62

QY 65 ISILEPKNLLFYH-----LISLKNQREGCSTD-----NAVSTPEFNR 104
 Db 63 ---LEFNTHPYVADPIIGVTLALBEEBANGADIDDAIQGVKVMGPIAGIGDVF-- 117
 QY 105 NLWY-----RLSSRFSIMKSCPRF-----FLDYLEAF----- 133
 Db 118 --MFTVRPIVGAIAASLAGSGSIAPLFFPIVMAAIRIAFIWYQEFYKSGSAITKDG 175
 QY 134 -GLISDFLDQAVIKFPELETHPSYVSGFVAPHOYS---TLQDREY--FPJASVMTL 187
 Db 176 GGLIQGTATKASITLGMFVGLIQRWVTINFGNPAVYSKIPLOKGAYLEFPFGSVSGT- 234
 QY 188 DKONPSLTPLDILHGLGHPWMLHPSFSSEFFINMGRLEFTKVEIKVQALPSKQRIQTLS 247
 Db 235 -----QMDILQVGS-----NKLSDPFR--VYVLD 259
 QY 248 NLIAIV-----RCFW 257
 Db 260 NLNDLIPGLAGLITPLCWM 279

RESULT 41

ABG15643
 ID ABG15643 standard; Protein; 744 AA.

XX AC ABG15643;
 XX 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #15634.
 XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001MO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS79830.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 XX PT diagnostics, forensics, gene mapping, identification of mutations
 XX PT responsible for genetic disorders or other traits and to assess
 XX PT biodiversity -
 XX PS Claim 20; SEQ ID No 46002; 103bp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue; as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABQ0010-ABQ030177 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 744 AA;

Query Match 4.9%; Score 93; DB 22; Length 744;
 Best Local Similarity 19.2%; Pred. No. 3.2; Mismatches 152; Gaps 22;
 Matches 87; Conservative 73;
 QY 3 YCERLDPKYLKALKLQSLSLFQNSQSLORAYSTPYYY-RIILOKENEKQAL-- 59
 DB 335 YAPNTGAPRFKQVLSLQKDL-----DSHTLIMGDFNPLSLDRSMRQVKNKDTQELNS 390
 QY 60 ARKICISILEFFKVL-----LFVHLLS-----LSKNQR-----EGCSTD 93
 DB 391 ALHOA-PLIDIYRTLHPKSTAYTFFSAPHHTYSKIDHILGSKALLSKCKRTEIITNCLSD 449
 QY 94 MAVVSTPPFNRNLWYLLSSRFLSKSYCPREFLDYLEAFGLLSDFLDH---QAVIK-FF 149
 DB 450 HSAIKLELRKN-----LTQNRSTTKW-----LNNL-----LNDYVWVHNEKAEIKMFF 494
 QY 150 ELE-----THFSYYPVSGFVAPHQYLSL-----LDQRYFPIASVMRTLDKDNFSLTPD 197
 DB 495 ETNENKDTTYQNLWDAPKACVCRKFTALNAHKRQKQSKIDTLTSQLEKEQEN--- 550
 QY 198 LIHDLGLGHVPWLLHPSPSEFFINMGLFTKVEK-----VOALPSKKQRIQTLOSNIJAI 252
 DB 551 -----KIDPLARLIKKEKNQIDAINKDKGIDITDPTIKTT 588
 QY 253 VRCFWFTVESGLIENH-----GRKAYGAVLS-----S 281
 DB 589 IREYIKHLYANKLENEMDKFDKYLPRLNQBEVESLNRPITGAEIVAINSLPTKKS 648
 QY 282 P-----QELGHAFIDNVRLVPLEDQIIRLPFNSTPTQETL-----FSIRHFDVELTS 331
 DB 649 PGPDGFTAEFYQRYKELLVCPFTLFE--KHFPVNYDFPQETLHFSPLSRSSBELNSTY 706
 QY 332 KL-----EMMLDQGLLESIPLYNQ 350
 DB 707 RLYGWSEBQSLWIRTWARVQKPRAAVPLVNE 738

RESULT 42
 ID AAU96092
 AC AAU96092; Protein; 333 AA.

15-JUL-2002 (first entry)

RhCMV unique short region US28.4 protein.

XX CMV; US28; vaccine; virucide; neuroprotective;
 KW antiinflammatory; ophthalmological; retinitis; encephalitis;
 KW pneumocystis.

OS Rhesus cytomegalovirus.

XX WO200218954-A2.

XX 07-MAR-2002.

XX 30-AUG-2001; 2001WO-US27392.

XX 30-AUG-2000; 2000US-229365P.

XX (CHEM-) CHEMOCENTRYX INC.

PI Schall TJ, Penfold M;
 XX WPI; 2002-351718/38.
 DR N-PSDB; ABK68823.

XX Isolated or recombinant homologues of US28 proteins and nucleic acids
 PT encoding the proteins, for use in vaccine compositions for treating an
 PT animal infected with, or at risk of infection by, cytomegalovirus -

XX Claim 69; Page 67; 95pp; English.

XX The invention relates to an isolated or recombinant homologue of US28
 CC protein which binds a chemokine (encoded by an open reading frame in the
 CC unique short (US) region in human cytomegalovirus (CMV) genome, having
 CC at least 75% identity to a rhesus monkey RHUS28.1-RHUS28.5; RH78 and
 CC RHUS33 (encoded by an open reading frame 78 in the unique long (UL) and
 CC ORF 33 is the US region of CMV genome) and their encoding nucleic acids.
 CC Also included are a vector comprising the nucleic acids, a cell
 CC comprising the nucleic acids, a vaccine comprising an immunogenic CMV
 CC polypeptide encoded by at least a region of CMV genome in which the
 CC polynucleotide sequence encoding US28 or its homologue has been
 CC inactivated, identifying an agent that reduces CMV dissemination in an
 CC animal, by determining whether the agent inhibits the expression or
 CC activity of US28 or US28 homologue, or fragment or variant of US28 or
 CC US28 homologue and treating an animal infected with CMV or at risk of
 CC infection by CMV, by administering the agent to the animal. The vaccine
 CC is useful for treating an animal infected with cytomegalovirus or at a
 CC risk of infection by CMV. The vaccine is also useful for inducing a
 CC therapeutic or protective immune response in a patient and in methods for
 CC treating diseases including retinitis, encephalitis and pneumocystis
 CC caused by CMV infection. The present sequence represents a human CMV or
 CC monkey CMV US or UL protein used to construct the vaccine of the
 CC invention.

XX Sequence 333 AA;

Query Match 4.9%; Score 92; DB 23; Length 333;
 Best Local Similarity 21.7%; Pred. No. 1.3;
 Matches 60; Conservative 44; Mismatches 101; Indels 72; Gaps 12;

QY 61 RHKICISILEFPKLLFVHLLSLSKNQREGCSTDMAVVSTPPFNRNLWYLLSSRFLSKWS 120
 DB 50 RHRTNS---FSDVLPFRHLM-----ITBEVFTLTP-----VWAYHLTTHGNLPGS 91
 QY 121 YCPRF-FUDYLEAFGLLSDFLDHQAIVKFELETHFSYYPVSGFVAPHQYLSLQDRYFP 179
 DB 92 WCRSLTFVFLTVFAR-----AFF-----YLLLIWDYRSV 121
 QY 180 IASVMRTLDKDNFSLTPDLIHLGHVPWLLHP-SFSEFFINMGLFTKVIKQVLPSPK 238
 DB 122 IICRHPLPVNLNYS-----QVIGLSVWLVAVLSASPFSIFNGSV-KQCLGNMGSIPTSE 173
 QY 239 KQRIQTLOSNIJAIVRCFWFTVESGLIENHGRKAYGAVLSQPELGHAFIDNVRLVPL 298
 DB 174 SSAVLNLEVHLS-----FWLPLIM-----SANCYQAKRRASPDQLHELYRCSLLITII 223
 QY 299 ELDQIIRLPFNSTPTQETLFSIRHFDVELTSKLEW 335
 DB 224 TTYAIVMFPFHLALLIDALLISISH-----VEPSSALHW 256

RESULT 43
 ID AAU33268
 AC AAU33268; Protein; 919 AA.

18-DEC-2001 (first entry)

Novel human secreted protein #3759.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX Homo sapiens.
 OS
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552928.
 XX
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YF, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 PS
 XX Claim 20; Page 750; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SO Sequence 919 AA;
 Query Match 4.9%; Score 92; DB 22; Length 919;
 Best Local Similarity 20.7%; Pred. No. 5.5; Indels 68; Gaps 12;
 Matches 61; Conservative 48; Mismatches 118;
 QY 3 YCERLDPKYLKIALKRSLSLFPONSQSLQRAYSTPYSYR---IIQKENKEQAL 59
 DB 567 YTRBELCTMFIREGVTLQSVYSKVNGSEIL-----FSYQDLVITLPFE-----L 613
 QY 60 ARHKCISILEFFKRLLFVHLISLKNQREGCSTDAVAVSTPFNRNLW-----YRLISS 113
 DB 614 RKRLIDVISMYREL-----LKDLSKAQGVFAQISLKTTEVL-RNLDLLQPIQLIJD 668
 QY 114 RFSIMKSCPRFFLDYL--EAFGLSDPFDHQAIVKFELETHSYYPVSGFAP----- 166
 DB 669 NIKOLKEMKFTYLINVIDEINTIFNDYIPY--VFKLKKNLCLINHKNEFIQNELQEA 726
 QY 167 -----HOYLSLQDRYFPASVWRT-----LDKDNFSLTPDLI----- 199
 DB 727 SGLQGIHQYIMALEEYFDDPSIVGMVKKYELKEKIVSLIKNLVALNDFSEYIVSAS 786
 QY 200 ---HDLGHVPLWLPSPSEPFINMGRLFTFVIEKVOALSKKQRIQTOSNLIA 251
 DB 787 NFTSQSSQVBOGFLHRTQOEYSLITDPDGKGEKIKINELATQOEI--IKSQALIA 839
 RESULT 44
 AAU31237
 ID AAU31237 standard; Protein: 3923 AA.
 XX

AC AAU31237;
 XX
 DT 08-NOV-1999 (first entry)
 XX
 DE Human Apo B protein fragment.
 XX
 KW Apo B; hyperlipidemia; human; treatment; hepatocyte; apoprotein;
 KW Apo A1; low density lipoprotein; LDL; blood; therapy; atherosclerosis;
 KW high density lipoprotein; HDL; cholesterol; coronary heart disease;
 KW Alzheimer's disease; hypobetalipoproteinemia; dybetalipoproteinemia.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..3923
 FT Protein /label= Apo B
 FT /note= "partial sequence, no start codon given"
 XX
 PN WO9940789-A1.
 XX
 PD 19-AUG-1999.
 XX
 PF 28-AUG-1998; 98WO-US17908.
 XX
 PR 30-JUN-1998; 98US-0106006.
 PR 12-FEB-1998; 98US-0074497.
 XX
 PA (MINU) UNIV MINNESOTA.
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 PI Bandypadhyay PT, Kren BT, Roy-Chowdhury J, Steer CJ;
 XX
 DR WPI; 1999-527333/44.
 XX
 DR N-PSDB; AAZ09525.
 XX
 PT Mutating apolipoprotein genes in hepatocytes to control cholesterol
 PT levels, e.g. for treating or preventing hyperlipidemia, particularly
 PT atherosclerosis
 XX
 PS Claim 4; Page 75-83; 106pp; English.
 XX
 CC This invention describes a novel method for the genetic treatment of
 CC hyperlipidemia by altering genes, in hepatocytes, for apoprotein (apo) B,
 CC E or A1. Low density lipoprotein (LDL) levels in the blood are reduced by
 CC altering an apo B gene (1) in a hepatocyte. The invention describes a
 CC method for the therapeutic and/or prophylactic method involving altering
 CC an apo B gene in hepatocytes by introducing the mutations Arg12Cys,
 CC Arg156Cys or Cys158Arg and a method for ameliorating atherosclerosis by
 CC altering the apo A1 gene in a hepatocyte so that the altered protein can
 CC dimerize. Altering expression of apo genes regulates levels of high and
 CC low density lipoprotein cholesterol. Altering expression of apo B, E and
 CC A1 genes is used to treat or prevent atherosclerosis, coronary heart
 CC disease, Alzheimer's disease, hypobetalipoproteinemia, and
 CC dybetalipoproteinemia. This sequence represents a fragment of the human
 CC Apo B protein described in the method of the invention.
 XX
 SO Sequence 3923 AA;
 Query Match 4.9%; Score 92; DB 20; Length 3923;
 Best Local Similarity 20.7%; Pred. No. 44;
 Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;
 QY 3 YCERLDPKYLKIALKRSLSLFPONSQSLQRAYSTPYSYR---IIQKENKEQAL 59
 DB 3571 YTRBELCTMFIREGVTLQSVYSKVNGSEIL-----FSYQDLVITLPFE-----L 3617
 QY 60 ARHKCISILEFFKRLLFVHLISLKNQREGCSTDAVAVSTPFNRNLW-----YRLISS 113
 DB 3618 RKRLIDVISMYREL-----LKDLSKAQGVFAQISLKTTEVL-RNLDLLQPIQLIJD 3672
 QY 114 RFSIMKSCPRFFLDYL--EAFGLSDPFDHQAIVKFELETHSYYPVSGFAP----- 166
 DB 3673 NIKOLKEMKFTYLINVIDEINTIFNDYIPY--VFKLKKNLCLINHKNEFIQNELQEA 3730

QY 167 -----HOYLSLLODRYFPFIASVMRT-----LDKNFSLTPDLI----- 199
 DB 3731 SOELQOIHOYIMALREYFDPISVGTWTVKYVELEEKIVSLIKNLLVALKDFHSEYIVSAS 3790
 QY 200 ---HDLGHVPLWLPSPFSEFFINMGRLEFTKVIKVOALPSKKQRIOTLQSNLIA 251
 DB 3791 NFTSOLSSQVQFLHRNIQIYLSILTDPDGKGEKIEALSATAOEI--IKSQAIA 3843

RESULT 45
 AAW41262
 ID AAW41262 standard; peptide; 4536 AA.
 AC AAW41262;
 DT 19-MAY-1998 (first entry)
 DE Apolipoprotein B-100.
 KW Anti-coagulant; apolipoprotein B-100; apoB-100; metastatic spread;
 KW thromboplastin-mediated process; cancer; inhibitor; blood coagulation;
 KW angiogenesis; cellular differentiation; apoptosis; KRAB-14;
 KW prothrombinase complex.
 OS Homo sapiens.
 PN WO9743311-A1.
 PD 20-NOV-1997.
 PF 09-MAY-1997; 97WO-GB01255.
 PR 09-MAY-1996; 96GB-0009702.
 XX (UNLO) ROYAL FREE HOSPITAL SCHOOL MED.
 PI Bruckdorfer KR, Bettelaie C;
 DR WPI, 1998-008798/01.
 PT Peptide fragments of apolipo:protein B-100 with anticoagulant
 PT activity - used for treating or preventing coagulation, inhibiting
 PT angiogenesis, cell differentiation and apoptosis
 PS Disclosure; Page 42-47; 60pp; English.
 CC This sequence represents the human apolipoprotein B-100 (apoB-100).
 CC Fragments of this sequence can be used in the peptide of the invention,
 CC which has the formula (I), or their variants with one or more internal
 CC deletions, insertions or substitutions, while retaining anti-coagulant
 CC properties of apolipoprotein B-100 (apoB-100).
 CC Z1-KAQ-X1-KKNKRRHS-X2-T-22 (I); X1 = S or Y; X2 = T or I;
 CC Z1 = the N terminus of the peptide, or 1-47 amino acids (aa);
 CC Z2 = the C terminus of the peptide, a terminal amide group or 1-77 aa.
 CC Compositions containing the peptide are used for simultaneous, separate
 CC or sequential treatment of cancer, particularly to prevent metastatic
 CC spread. They are also used to inhibit thromboplastin-mediated processes,
 CC specifically to prevent or reduce blood coagulation (e.g. during or after
 CC surgery or in cases of heart attack, stroke etc.) and to inhibit
 CC angiogenesis, cellular differentiation or apoptosis. KRAB-14, which is
 CC active as such or as part of a 98-aa peptide, inhibits activation of the
 CC prothrombinase complex; and prevents activation of factor VII on the
 CC surface of thromboplastin and of platelets by thrombin. It binds to the
 CC residues 58-66 of thromboplastin. Since (I) are much smaller than
 CC apoB-100, they act more quickly.
 XX Sequence 4536 AA;
 SQ

QY 3 YCERLDPKYILKIALKRLQSLSLFFQNSQSLQRAYSTPYSYR---IILOKENKEKQAL 59
 DB 4184 YTBRELCTMFTREVGTGTVLSQVSKVHNGSEIL-----FSYFQDLVITLPFE-----L 4230
 QY 60 AHKHCISILEFFKNLLFVHLLSLSKNQREGCSTDMVVSTPFFNRNLW-----YRLSS 113
 DB 4231 RKHLIDVISMYREL-----LKDLSEKQAVFKAIOSLKTTTEVL--RNLQDLLQIFQIED 4285
 QY 114 RESLWKSYPFRFLDYL--EAFGLLSDFLDHQAVIKFFFELETHFSYVPSGFWAP----- 166
 DB 4286 NIKQLKEMKFTYLYNIYIQDEINTIFNDYIPY--VFKLKENLCLNLHKHNEFIQNELQEA 4343
 QY 167 -----HOYLSLLODRYFPFIASVMRT-----LDKNFSLTPDLI----- 199
 DB 4344 SOELQOIHOYIMALREYFDPISVGTWTVKYVELEEKIVSLIKNLLVALKDFHSEYIVSAS 4403
 QY 200 ---HDLGHVPLWLPSPFSEFFINMGRLEFTKVIKVOALPSKKQRIOTLQSNLIA 251
 DB 4404 NFTSOLSSQVQFLHRNIQIYLSILTDPDGKGEKIEALSATAOEI--IKSQAIA 4456

RESULT 46
 AAW96826
 ID AAW96826 standard; protein; 4536 AA.
 AC AAW96826;
 DT 22-APR-1999 (first entry)
 DE Amino acid sequence of human apolipoprotein B-100 (apoB-100).
 KW Human apolipoprotein B-100; apoB-100; very-low density lipoprotein; VLDL;
 KW apolipoprotein; binding; in vivo transport; nucleic acid; binding domain;
 KW nuclear localisation sequence; gene therapy; cancer; cystic fibrosis;
 KW non-small cell lung carcinoma; diabetes; arteriosclerosis.
 OS Homo sapiens.
 PN WO9856938-A1.
 PD 17-DEC-1998.
 PF 10-JUN-1998; 98WO-US11927.
 PR 14-MAY-1998; 98US-0079030.
 PR 13-JUN-1997; 97US-0874807.
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 XX Guevara JG, Hoogveen RC, Moore JP;
 WPI; 1999-070331/06.
 CC Composition comprising nucleic acid bound to LDL or VLDL lipoprotein
 CC - used for delivering nucleic acid to cells for gene therapy and
 CC antisense treatment
 PS Disclosure; Fig 1A-C; 293pp; English.
 CC The present sequence represents human apolipoprotein B-100 (apoB-100).
 CC ApoB-100 is a major apoprotein component of very-low density
 CC lipoproteins (VLDL), intermediate density lipoprotein (IDL), low
 CC density lipoproteins (LDL) and lipoprotein a. The specification
 CC describes a composition that comprises LDL and apolipoproteins for the
 CC binding and in vivo transport of nucleic acids. Binding domains
 CC (see AAW96827-77) and nuclear localisation sequences (see AAW96878-97)
 CC for use in the composition can be derived from the present sequence. The
 CC composition is used to deliver nucleic acids to eukaryotic cells, in
 CC vivo or in vitro, for expressing a therapeutic polypeptide or antisense
 CC molecule (or ribozyme). Specifically they are used for gene therapy of
 CC cancers (particularly non-small cell lung carcinoma), diabetes, cystic
 CC fibrosis and arteriosclerosis.
 XX

QY 167 -----HOYLSLLODRYFPFIASVMRT-----LDKNFSLTPDLI----- 199
 DB 3731 SOELQOIHOYIMALREYFDPISVGTWTVKYVELEEKIVSLIKNLLVALKDFHSEYIVSAS 3790
 QY 200 ---HDLGHVPLWLPSPFSEFFINMGRLEFTKVIKVOALPSKKQRIOTLQSNLIA 251
 DB 3791 NFTSOLSSQVQFLHRNIQIYLSILTDPDGKGEKIEALSATAOEI--IKSQAIA 3843

Query Match 4.9%; Score 92; DB 19; Length 4536;
 Best Local Similarity 20.7%; Pred. No. 55;
 Matches 61; Conservative 48; Mismatches 118; Indels 68; Gaps 12;

XX WPI; 1999-562049/47.
 DR N-PSDB; AAZ19859.
 XX
 PT New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used
 PT for DNA amplification, sequencing, polymerase chain reaction and
 PT reverse transcription
 XX
 PS Example 6; Fig 2; 71pp; English.
 XX
 CC This sequence represents the DNA polymerase of Thermoanaerobacter
 CC thermohydrosulfuricus (Tt), deduced from PCR amplified genomic
 CC DNA (see AAZ19859). Claimed DNA polymerases (see AAY31813, AAY31815 and
 CC AAY31816) have the exonuclease activity of this enzyme removed by an
 CC having 540-582 amino acids removed from the N-terminus or having an
 CC amino acid substitution at position 8 or 706 of the sequence. A
 CC novel method for PCR uses an enzymatically active DNA polymerase
 CC that has at least 80% identity in its amino acid sequence to the Tt
 CC DNA polymerase and has an exonuclease activity removed. Kits for
 CC PCR, strand displacement amplification of DNA, DNA sequencing,
 CC and reverse transcription-PCR are claimed that utilize such Exo-
 CC DNA polymerases. The Tt DNA polymerase is stable at elevated
 CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity
 CC does not affect its ability to replicate DNA or its thermostability.
 XX
 SQ Sequence 872 AA;
 Query Match 4.8%; Score 91.5; DB 20; Length 872;
 Best Local Similarity 17.5%; Pred. No. 5.7;
 Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;

QY 14 LKIALKQSLPFFONSQSLORAYSTPYVYRILOKENKEKQALAHKICISILEFFKN 73
 DB 224 IKENIENNKELAINSKELATIKEDIPIDFEEYKVKFNEEK-----LLELPNK 273
 QY 74 LLFVHLLSLKNQREGCGTDMVSTPFFNRNLWYLLSSRFLSKSCYCRFPFLDYLEAF 133
 DB 274 LEFSLDINIKKE-----SSIEVD-----NHKVERKSK-----VDIKELV 309
 QY 134 GLLSDFLDHQAVIKFELETHFSYYPV-----SGFVAPHQYLSLLQDRYFFIASV 183
 DB 310 TLLQD-----NNNIAFYELIYEGEIKKIAFSFGKOTVYIDVFQTE-----DL 351
 QY 184 MRTLKDNFSLTPDLIHDL-----GHVPWLLHPSFSFFINMGRLEFT 226
 DB 352 KEIFEKEDFEFTTHEIKDFLVLRSYKGIKESKYIDTAVMAYLLNPSESNY--DLDRVLK 409
 QY 227 KVIEKVOALPS-----KKORIQTLOSNLIA-----IVRCFW----- 257
 DB 410 KYL-KVD-VPSYEGIFGKGRDKKKIEIDENILADYICSRGVYFLDLKELMNFIEMDM 467
 QY 258 -----FTVESGIEHNHGRKAYGAVLSSP-QELGHAFIDNVRVLPFLDQIIR---LPFN 309
 DB 468 KKLLEIEMLPVELVKMEVSGPTLDEKVELKSKIDDRIGEI---LDKIYKAGVQFN 524
 QY 310 TSTPQE-TLFSIRHFDLVELTSLKLEWLDQGLLESIPLYN-----QEKY 353
 DB 525 VNSPKQLSEFLFKLNLPIVKTKTGYSTDSEVLQVLPYNDIIVRYQLTKLSTY 584
 QY 354 LSGF 357
 DB 585 IDGF 588

RESULT 49
 AAY31815
 ID AAY31815 standard; Protein; 872 AA.
 XX
 AC AAY31815;
 XX
 DT 06-DEC-1999 (first entry)
 XX
 DE Thermoanaerobacter thermohydrosulfuricus DNA polymerase D8A mutant.

XX DNA polymerase; thermostable enzyme; amplification; sequencing;
 KW PCR; reverse transcription; exonuclease; mutant.
 XX
 OS Thermoanaerobacter thermohydrosulfuricus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 8 /note= "wild-type Asp is substituted by Ala"
 XX
 PN WO9947539-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 17-MAR-1999; 99WO-US05612.
 XX
 PR 18-MAR-1998; 98US-0044106.
 XX
 PA (AMSH) AMERSHAM PHARMACIA BIOTECH INC.
 XX
 PI Mamone JA, Davis M, Sha D;
 XX WPI; 1999-562049/47.
 XX
 PT New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used
 PT for DNA amplification, sequencing, polymerase chain reaction and
 PT reverse transcription
 XX
 PS Claim 5; Page 7; 71pp; English.
 XX
 CC This sequence represents a claimed mutant DNA polymerase of
 CC Thermoanaerobacter thermohydrosulfuricus (Tt) in which the Asp-8
 CC residue of the native enzyme is replaced by Ala. This amino acid
 CC substitution removes the exonuclease activity of the DNA
 CC polymerase. Claimed DNA polymerases have their exonuclease
 CC activity removed by N-terminal deletion or amino acid substitution.
 CC A novel method for PCR uses an enzymatically active DNA polymerase
 CC that has at least 80% identity in its amino acid sequence to the Tt
 CC DNA polymerase and has an exonuclease activity removed. Kits for
 CC PCR, strand displacement amplification of DNA, DNA sequencing,
 CC and reverse transcription-PCR are claimed that utilize such Exo-
 CC DNA polymerases. The Tt DNA polymerase is stable at elevated
 CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity
 CC does not affect its ability to replicate DNA or its thermostability.
 CC Note: the present sequence is not shown in the specification but is
 CC derived from the Tt DNA polymerase sequence given in figure 1.
 XX
 SQ Sequence 872 AA;

Query Match 4.8%; Score 91.5; DB 20; Length 872;
 Best Local Similarity 17.5%; Pred. No. 5.7;
 Matches 74; Conservative 70; Mismatches 141; Indels 139; Gaps 19;
 QY 14 LKIALKQSLPFFONSQSLORAYSTPYVYRILOKENKEKQALAHKICISILEFFKN 73
 DB 224 IKENIENNKELAINSKELATIKEDIPIDFEEYKVKFNEEK-----LLELPNK 273
 QY 74 LLFVHLLSLKNQREGCGTDMVSTPFFNRNLWYLLSSRFLSKSCYCRFPFLDYLEAF 133
 DB 274 LEFSLDINIKKE-----SSIEVD-----NHKVERKSK-----VDIKELV 309
 QY 134 GLLSDFLDHQAVIKFELETHFSYYPV-----SGFVAPHQYLSLLQDRYFFIASV 183
 DB 310 TLLQD-----NNNIAFYELIYEGEIKKIAFSFGKOTVYIDVFQTE-----DL 351
 QY 184 MRTLKDNFSLTPDLIHDL-----GHVPWLLHPSFSFFINMGRLEFT 226
 DB 352 KEIFEKEDFEFTTHEIKDFLVLRSYKGIKESKYIDTAVMAYLLNPSESNY--DLDRVLK 409
 QY 227 KVIEKVOALPS-----KKORIQTLOSNLIA-----IVRCFW----- 257
 DB 410 KYL-KVD-VPSYEGIFGKGRDKKKIEIDENILADYICSRGVYFLDLKELMNFIEMDM 467

Fri Jan 10 10:56:10 2003

us-09-438-185a-1047.rag

Page 34

QY 258 ----FTVESGLIENHEGRKAYGAVLISSP-QELGHAFIDNVAVLPLELDQIIR---LPPN 309
DB 468 KLLLEIEMPLEVLEKSMESVGFILDKVEVLELSQKIDRIGEI---LDKIYKAGYQFN 524
QY 310 TSTPOE-TLFSIRHPELVELTSLKLEMLDQGLLESIPLYN-----OEKY 353
DB 525 VNSPKQSEFLPEKLNLPVKKTKGTGSTDEVLQVLPYNDIVSDIIEYRLTKLSTY 584
QY 354 LSGP 357
DB 585 IDGF 588
RESULT 50
AAV31816
ID AAV31816 standard; Protein; 872 AA.
XX
AC AAV31816;
XX
DT 06-DEC-1999 (first entry)
XX
DE T. thermohydrosulfuricus DNA polymerase F706Y mutant.
XX
KM DNA polymerase; thermostable enzyme; amplification; sequencing;
XX PCR; reverse transcription; exonuclease; mutant.
XX
OS Thermoanaerobacter thermohydrosulfuricus.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 706
FT /note= "wild-type phe is substituted by Tyr"
XX
PN WO9947539-A1.
XX
PD 23-SEP-1999.
XX
PF 17-MAR-1999; 99WO-US05612.
XX
PR 18-MAR-1998; 98US-0044106.
XX
PA (AMSH) AMERSHAM PHARMACIA BIOTECH INC.
XX
PI Mamone JA, Davis M, Sha D;
XX WPI; 1999-562049/47.
DR
XX
XX
PT New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus, used
PT for DNA amplification, sequencing, polymerase chain reaction and
PT reverse transcription -
XX
PS Claim 23; Page -; 71pp; English.
XX
CC This sequence represents a claimed mutant DNA polymerase of
CC Thermoanaerobacter thermohydrosulfuricus (Tt) in which the Phe-706
CC residue of the native enzyme is replaced by Tyr. This amino acid
CC substitution removes the exonuclease activity of the DNA
CC polymerase. Claimed DNA polymerases have their exonuclease
CC activity removed by N-terminal deletion or amino acid substitution.
CC A novel method for PCR uses an enzymatically active DNA polymerase
CC that has at least 80% identity in its amino acid sequence to the Tt
CC DNA polymerase and has an exonuclease activity removed. Kits for
CC PCR, strand displacement amplification of DNA, DNA sequencing,
CC and reverse transcription PCR are claimed that utilize such exo-
CC DNA polymerases. The Tt DNA polymerase is stable at elevated
CC temperatures, e.g. 70+ deg C. Removal of its exonuclease activity
CC does not affect its ability to replicate DNA or its thermostability.
CC Note: the present sequence is not shown in the specification but is
CC derived from the Tt DNA polymerase sequence given in figure 1.
SQ Sequence 872 AA;

Query Match 4.8%; Score 91.5; DB 20; Length 872;
Best Local Similarity 17.5%; Pred. No. 5.7; 141; Indels 139; Gaps 19;
Matches 74; Conservative 70; Mismatches 139; Gaps 19;
QY 14 LKTAIKRQSLSPFQNSQLOKAYSTPYRIIQQEKKEKQALARKKISIIIEFFN 73
DB 224 IKENIENKRLAISMGLATIGDPIIDEEYKKNFEK-----LLELFNK 273
QY 74 LLPVHLISLKNQRCGCTDNAVSTPPFNMLYRLLSRFSLKSYCPRFFLDYLA 133
DB 274 LEFSLIDNIKE-----SSIEIVD-----NHKVERKSK-----VDIKELV 309
QY 134 GLISDFLDHQAVIKFELETHFSYYPV-----SGFVAPHOYLISLQDRYFPIASV 183
DB 310 TLIDP-----NRNIAFPYPLIYEGEIKKIAFSFGKDTVIYIDVOTE-----DL 351
QY 184 MRTLDKDNFSLTPDLIHDL-----GHVPMLHPSFSEFFINMGRLEPT 226
DB 352 KEIPEKEDFEFTTHEIKDFLVRLSYKGECKXYIDTAVAVYLLNPESNY--DLDRVLK 409
QY 227 KVIKYOALPS-----KKORIQTQSLNLR--IVRCFM----- 257
DB 410 KYL-KVD-VPSYSGIRGKGRDKKIEIDENIADYICSCCYLFDLKEKLNFIEMDM 467
QY 258 ----FTVESGLIENHEGRKAYGAVLISSP-QELGHAFIDNVAVLPLELDQIIR---LPPN 309
DB 468 KLLLEIEMPLEVLEKSMESVGFILDKVEVLELSQKIDRIGEI---LDKIYKAGYQFN 524
QY 310 TSTPOE-TLFSIRHPELVELTSLKLEMLDQGLLESIPLYN-----OEKY 353
DB 525 VNSPKQSEFLPEKLNLPVKKTKGTGSTDEVLQVLPYNDIVSDIIEYRLTKLSTY 584
QY 354 LSGP 357
DB 585 IDGF 588

Search completed: January 9, 2003, 15:27:08
Job time : 77 secs